UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

Rafael Rubiati Scalvenzi

Classificação inteligente de sinais musicais utilizando a Transformada *Wavelet-Packet*

São José do Rio Preto
2018

Rafael Rubiati Scalvenzi

Classificação inteligente de sinais musicais utilizando a Transformada *Wavelet-Packet*

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Orientador: Prof. Dr. Norian Marranguello
Coorientador: Prof. Dr. Rodrigo Capobianco Guido

São José do Rio Preto
2018

Scalvenzi, Rafael Rubiati.
Classificação inteligente de sinais musicais utilizando a Transformada Wavelet-Packet / Rafael Rubiati Scalvenzi. -- São José do Rio Preto, 2018
107 p. : il. ; tabs.

Orientador: Norian Marranguello
Coorientador: Rodrigo Capobianco Guido
Dissertação (Mestrado) – Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas

1. Computação - Matemática. 2. Processamento de sinais – Técnicas digitais. 3. Autocorrelação. 4. Redes neurais (Computação) 5. Música e tecnologia. I. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas. II. Título.

CDU – 518.72

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Câmpus de São José do Rio Preto

Rafael Rubiati Scalvenzi

Classificação inteligente de sinais musicais utilizando a Transformada *Wavelet-Packet*

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Comissão Examinadora

Prof. Dr. Rodrigo Capobianco Guido
UNESP – Câmpus de São José do Rio Preto
Coorientador

Prof. Dr. Aledir Silveira Pereira
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Everthon Silva Fonseca
IFSP – Câmpus de Catanduva

São José do Rio Preto
20 de julho de 2018

AGRADECIMENTOS

Primeiramente a Deus, pois sem Ele nada é possível. Ele nos dá força, perseverança e nos mostra o caminho quando tudo parece impossível.

Ao Prof. Dr. Norian Marranghello, pelo incentivo e orientações, mostrando possibilidades, falhas e contribuindo muito para a realização deste trabalho.

Ao Prof. Dr. Rodrigo C. Guido pelo apoio, correções e orientações que levaram aos resultados obtidos neste trabalho.

Ao Prof. Jonathan G. Rogeri, pelo incentivo e colaboração ao longo dessa jornada.

À minha esposa Lêda, pelo apoio e motivação nos momentos mais difíceis. E a todos os meus amigos e companheiros, pelo seu apoio e amizade.

RESUMO

A área na qual a música está inserida requer, para sua compreensão, considerável abstração. Neste âmbito, a análise matemático-computacional possui papel importante, principalmente para planejar a interatividade entre aluno e computador, potencializando o aprendizado musical. Embora um número considerável de estudos em diferentes contextos sejam dedicados à classificação das estruturas sonoras, os procedimentos de análise em um grande conjunto de sinais podem tornar-se uma tarefa difícil e exaustiva. Diante do exposto, este trabalho tem como objetivo a proposição e a implementação de um método capaz de reconhecer e classificar sinais musicais em tempo real, visando auxiliar os aprendizes. No método proposto, um conjunto relevante de eventos musicais é inspecionado por meio da análise de multirresolução baseada na Transformada *Wavelet-Packet*, escolhida em função da característica multidimensional encontrada na música, a qual permite isolar diferentes eventos musicais em níveis de decomposição *wavelet* distintos. Apoiado por um processo de autocorrelação e uma rede neural artificial, cada padrão sônico é associado ao seu respectivo evento musical. Testes envolvendo centenas de sinais permitiram obter uma acurácia quase plena com um tempo relativamente bastante pequeno de análise em função da baixa ordem de complexidade computacional do algoritmo implementado, reafirmando a sua aplicabilidade.

Palavras-chave: processamento de sinais digitais, sinais musicais, transformada *wavelet-packet*, autocorrelação, redes neurais artificiais.

ABSTRACT

Music belongs to an area which requires a considerable piece of abstraction for its understanding. In this domain, computational and mathematical analyses play an important role, particularly for planning human-machine interaction and enhancing learning. Although a considerable number of studies in different musical contexts are dedicated to the classification of the structures present in sound signals, the inspection of long clips is a challenge. Thus, this work proposes and implements a method capable of identifying and classifying musical signals in real-time, helping music students. Specifically, multiresolution analysis using the Wavelet-Packet Transform is adopted, allowing for different musical events to be isolated in distinct wavelet levels of decomposition. Based on an autocorrelation and an artificial neural network, each sonic pattern is associated with a respective musical event. Tests using hundreds of music clips exhibit almost full accuracy with relatively very short time consumption as a function of the algorithm low level of computational complexity, reassuring its applicability.

Keywords: digital signal processing, musical signals, wavelet-packet transform, autocorrelation, artificial neural networks.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

MIDI	Musical Instrument Digital Interface
Hz	Hertz
ZCR	Zero Crossing Rate
AC	Autocorrelação
FS	Fourier Series
FT	Fourier Transform
DTFS	Discrete Time Fourier Series
DTFT	Discrete Time Fourier Transform
STFT	Short-Time Fourier Transform
CQT	Constant-Q Transform
AMR	Análise Multi-Resolucional
WT	Wavelet Transform
CWT	Continuos Wavelet Transform
DWT	Discrete Wavelet Transform
IDWT	Inverse Discrete Transform Wavelet
QMF	Quadrature Filter Mirror Pair
WPT	Wavelet Packet Transform
RNA	Rede Neural Artificial
MLP	Multi-Layer Percetron
RBF	Radial-Basis Function
VC	Vetor Característica

SUMÁRIO

1 INTRODUÇÃO

A tecnologia aplicada às mais diversas áreas vêm transformando o contexto de ensino e de aprendizagem, despertando o interesse nas pessoas em todas as áreas de conhecimento. A utilização da computação na educação musical é uma consequência natural da crescente demanda por recursos tecnológicos no fazer musical. Assim, a facilidade de interação entre aluno e computador pode vir a potencializar o aprendizado musical e expandir o universo de possibilidades musicais (MENESES; FORNARI, 2015).

Recentemente, a computação musical ganhou forte impulso com os ambientes para interação aluno-computador. O conceito de educação musical interativa é um assunto bastante novo visto que, até poucos anos, os intérpretes tinham que se adaptar a um material gravado, sincronizando seu desempenho e ajustando a dinâmica e qualidade sonora aos sons gravados (IAZZETTA, 1998). Por interatividade entende-se a possibilidade de acesso em tempo real a informação, fortalecendo a relação usuário-tempo-informação e refletindo diretamente nos processos de percepção. Assim, ampliam-se os sentidos humanos e estimula-se a capacidade de processar diversas informações simultaneamente (BARRETO, 1997). A interatividade em tempo real contribui diretamente no processo de ensino e aprendizagem, pois o aprendiz passa a assumir e atuar na construção do conhecimento, deixando de ser um mero expectador do processo.

Muitos músicos têm um bom apoio para compreender a formulação teórica da música, devido à grande disponibilidade de material didático que pode ser encontrado em livros e na Internet. Porém, o início da parte prática pode ser uma etapa de difícil execução, pois o aprendiz talvez ainda não tenha uma percepção musical boa, ou seja, não consegue identificar intuitivamente os eventos musicais básicos, como a altura das notas musicais ou, mesmo, o andamento de uma melodia. Ao reproduzir um trecho musical, o resultado acaba sendo, na maioria das vezes, uma melodia fora de tonalidade e sem ritmo. O domínio tanto teórico quanto prático é determinante na formação de um bom músico.

Diante do exposto, a motivação para o desenvolvimento deste trabalho está na

implementação de um método para reconhecimento de sinais musicais em tempo real, capaz de auxiliar na aprendizagem e de ajudar músicos iniciantes na atuação prática. Muitas pesquisas vêm sendo desenvolvidas no ramo da aprendizagem musical, avançando no campo de recursos virtuais, a exemplo dos ambientes para ensino a distância ou *softwares* na área de percepção musical (LEME; BELLOCHIO, 2014). Conta-se hoje com uma grande quantidade de trabalhos focados no processamento de sinais sonoros, partindo-se de diferentes contextos, entre os quais podem-se citar o reconhecimento de acordes musicais de violão utilizando a transformada de Fourier (FERREIRA, 2006), a utilização de *wavelets* para transcrição de sinais musicais (TREVILLATO; BARBEDO; LOPES, 2005), a classificação de gêneros musicais (GRIMALDI; CUNNINGHAM; KOKARAM, 2003), a extração de descritores sonoros timbrísticos (ROQUE; MENDES, 2014) e tempo musical (JR; DAMIANI, 2014), entre muitos outros.

A classificação das principais estruturas musicais presentes em um sinal de áudio tem sido uma tarefa de considerável dificuldade. Os problemas correspondentes à análise da percepção musical por computador podem ser resumidos na detecção do ritmo, da altura das notas e do timbre de instrumentos (NAGARAJ; EVANS, 2003). Nesse contexto, as técnicas de processamento de sinais, tal como a Transformada de *Wavelet*, aplicam-se em um vasto campo de modelamento e representação da música. A característica multidimensional da música, composta por estruturas sonoras de diferentes intensidades e durações, permite uma análise multirresolução com *wavelets* capaz de visualizar o sinal musical em diferentes níveis de resolução e isolar eventos musicais em níveis *wavelet* distintos (FARIA; ZUFFO, 1995). O emprego de *wavelets* na análise de sinais musicais oferece vantagens em comparação às outras técnicas, devido a sua baixa complexidade algorítmica e capacidade de localizar características e propriedades nos domínios temporal e espectral, implicando em uma ferramenta poderosa para aplicações musicais de tempo real.

1.1 Objetivos

Este trabalho tem por objetivo desenvolver um método capaz de reconhecer e classificar sinais musicais em tempo real, utilizando a Transformada *Wavelet-Packet*, para auxiliar alunos no aprendizado musical. Espera-se, além disso, contribuir com as

pesquisas na área da computação musical, auxiliando outros estudos que envolvam a percepção computacional da música por meio de processamento de sinais. Assim, organizou-se uma metodologia que procurou satisfazer os seguintes objetivos:

- realizar a análise de trechos musicais em tempo real, utilizando a Transformada *Wavelet-Packet* como técnica fundamental;
- identificar os eventos musicais codificados nas sub-bandas *wavelet* por meio de uma rede neural artificial;
- avaliar os resultados visando o auxílio no aprendizado musical.

1.2 Estrutura do trabalho

Este trabalho está organizado da seguinte forma. No Capítulo 2, abordam-se os conceitos fundamentais da música, teoria e representação dos sinais sonoros. Prosseguindo, o Capítulo 3 contém um resumo das principais técnicas utilizadas na área de análise de sinais, focalizando, em particular, aquelas que são adequadas à análise acústica, incluindo a Transformada *Wavelet-Packet* Discreta e as Redes Neurais Artificiais. Em seguida, a metodologia e o modo como foram empregados os métodos para realização do trabalho constam no Capítulo 4. Finalmente, o Capítulo 5 contém uma descrição dos resultados, e o Capítulo 6 é dedicado às conclusões, as quais são seguidas das referências.

2 REVISÃO BIBLIOGRÁFICA - PARTE I: MÚSICA E REPRESENTAÇÃO DIGITAL DE SINAIS

A área na qual a música está inserida exige grande abstração para plena compreensão, pois ela encontra-se apoiada na arte, na expressão de sentimentos e na ciência, obedecendo a leis físicas e universais. Os sinais sonoros permitem que nossa percepção auditiva adquira informações, e nosso cérebro perceba o ambiente apreciando assim a música e viabilizando a comunicação falada. Neste capítulo, serão descritos os conceitos básicos da teoria musical e dos sinais sonoros, seus termos e definições necessários para o desenvolvimento deste trabalho.

2.1 Música

A música é a arte da combinação dos sons, para os quais se deseja produzir uma sensação agradável aos nossos ouvidos. A palavra música, do grego *musiké téchne* (arte das musas), existe desde as primeiras civilizações, com manifestações próprias e teorias complexas. No século VI, Pitágoras demonstrou as primeiras conexões entre música e aritmética, estabelecendo relações matemáticas entre as frequências das notas da escala maior. A escola Pitagórica também propôs que combinações sonoras percebidas como agradáveis deveriam ser produzidas por instrumentos cujas dimensões estivessem relacionadas por frações simples (CARVALHO et al., 2009). Leonardo da Vinci fez uma série de aperfeiçoamentos nos instrumentos músicais da sua época e o compositor Johann Sebastian Bach explorou, na sequência, a possibilidade de usar todas as tonalidades numa só composição por meio do sistema temperado, resultando na sua obra genial “Cravo Bem Temperado” (CUNHA; MARTINS, 1998).

A partir do século XX, com as contribuições de diversas áreas da ciência tais como a psicologia, foi desenvolvida a Teoria Geral Musical, estruturada por meio de diversas disciplinas, como o solfejo, o ritmo, a percepção melódica, entre outras, tidas como um meio para o seu entendimento (JUNIOR, 2011). A estrutura musical determina que o material sonoro é tradicionalmente composto por três elementos fundamentais: a melodia, a harmonia e o ritmo. A melodia consiste na sucessão dos sons formando um sentido. A harmonia é a execução de vários sons ouvidos ao mesmo tempo, observadas as regras que regem sons simultâneos. O ritmo, por fim, é o movimento de

sons regulados pela sua duração (PRIOLLI, 2015).

A música pode ser representada de muitas formas diferentes, como por exemplo, uma partitura musical, na qual são utilizados símbolos chamados de notas, tocadas por um músico, ou protocolos padrão como o *Musical Instrument Digital Interface* (MIDI), utilizado em instrumentos eletrônicos onde as mensagens de evento especificam intensidade, velocidade e outros parâmetros para gerar os sons pretendidos (MÜLLER, 2015). Podemos distinguir três classes principais para representar a música: símbolos, partituras, e áudio. Os símbolos compreendem qualquer tipo de representação como uma codificação de notas ou eventos musicais. O termo partitura refere-se às representações visuais por meio de símbolos que se associam aos sons. O áudio, finalmente, refere-se à representação com base em ondas acústicas (MÜLLER, 2015).

2.1.1 Melodia, harmonia e ritmo

Dada a natureza artística da música, a definição precisa dos conceitos de melodia e harmonia é impossível de um modo que não inclua outras características, como ritmo e percussão. De modo mais simples, pode-se afirmar que melodia e harmonia são características musicais atreladas à noção de nota musical (CARVALHO et al., 2009). A característica que define uma melodia como uma sucessão de sons que fazem sentido, ou seja, que sejam agradáveis aos nossos ouvidos, está inteiramente ligada às propriedades do ritmo e da harmonia.

Na música, a harmonia refere-se à simultaneidade de sons diferentes, ou seja, quando duas ou mais notas são tocadas ou ouvidas ao mesmo tempo. Ela pertence à classe de sons do tipo polifônicos, podendo conter várias notas tocadas simultaneamente, ao contrário dos sons monofônicos, em que há apenas uma nota presente. Os principais componentes constitutivos da harmonia, pelo menos na tradição da música ocidental, são os acordes, isto é, construções musicais que tipicamente consistem em três ou mais notas. A análise de harmonia pode ser pensada como o estudo da construção, interação e progressão de acordes (MÜLLER, 2015).

Os fenômenos que emergem da distribuição de sons no tempo podem ser vistos como fenômenos rítmicos (CARVALHO et al., 2009). Adotando a notação musical clássica, observa-se que algumas notas de frases musicais são mais acentuadas

que outras, muitas vezes periodicamente, e que essas notas são quase sempre a primeira de cada compasso. A duração de cada nota é normalmente uma subdivisão em partes iguais do período de acentuação, criando uma estrutura hierárquica de níveis de subdivisão chamada de estrutura métrica. Ela geralmente é medida em pulsos (batidas), tal como a frequência em que batemos o pé ao ouvirmos música, ou a frequência com que o maestro move a batuta em uma regência. Esses pulsos na música clássica são geralmente representados por batidas por minuto (bpm).

As combinações de pulsos temporais podem resultar em infinitas variações de ritmo. Na música, essas variações têm uma só derivação nos dois ritmos fundamentais, chamados de ritmo binário e ritmo ternário. O ritmo binário corresponde à divisão de uma unidade de tempo em duas partes iguais e o ternário, à divisão de uma unidade de tempo em três partes iguais. Por unidade de tempo entende-se o espaço de tempo que se passa entre dois limites preestabelecidos e sensíveis ao ouvido (POZZOLI, 1983). Os ritmos binários são formados pela sucessão de um acento forte e de um fraco, enquanto que o ritmo ternário, pela sucessão de um acento forte e dois fracos. Entende-se por acentuação forte, meio forte ou fraca os momentos ou tempos que não dão a mesma impressão acústica.

2.2 A Escala Temperada

A partir do século XVIII, foi introduzida por meio de Bach uma padronização para sons produzidos por instrumentos musicais, adotando-se uma afinação padrão, gerando a chamada Escala Bem Temperada (LIMA, 2006). Ela é a sucessão de oito sons conjuntos, ou oito graus, chamados de oitava, onde tem-se o oitavo grau como a repetição do primeiro. A cada oitava, os sons se reproduzem novamente, soando similares aos nossos ouvidos, devido ao fato de que, em cada oitava na escala bem temperada, as notas do mesmo nome possuem frequências múltiplas da escala anterior (OLIVEIRA, 2007a).

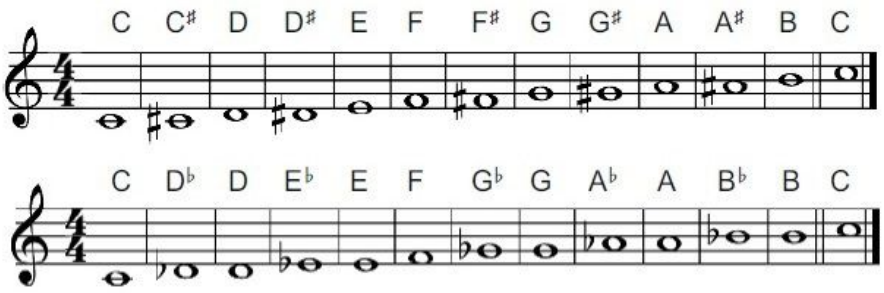
Os intervalos idênticos entre as notas das escalas são definidos por meio de uma progressão geométrica de frequências, sempre com razão constante e igual a $\frac{2}{12}$ para notas consecutivas. Em princípio, tem-se sete notas musicais, sendo elas dó, ré, mi, fá, sol, lá e si, também podendo ser representadas respectivamente pelas cifras

C, D, E, F, G, A e B. Estas sete notas podem adquirir a condição de sustenido(#) ou bemol(b), formando os chamados semitons. De acordo com o sistema temperado, temos então exatamente doze notas consecutivas em cada oitava, cada uma soando diferente da outra.

Um semitom é o menor intervalo utilizado na música ocidental, correspondendo, por exemplo, à diferença de altura ascendente ou descendente entre duas teclas adjacentes do piano (uma branca e uma preta adjacentes), com duas exceções: entre Mi e Fá, e entre as notas Si e Dó (no piano, são duas teclas brancas adjacentes). O tom é definido como a soma da distância ascendente ou descendente de dois semitons.

A escala musical pode ser considerada como um conjunto de notas no qual os elementos geralmente são ordenados por passos ascendentes. Cada passo da escala é um intervalo entre duas notas sucessivas. Enquanto um acorde pode ser pensado como uma estrutura vertical, as escalas são geralmente associadas a estruturas horizontais. Assumindo o princípio da equivalência de oitava (MÜLLER, 2015), as escalas geralmente ocupam uma única oitava, com oitavas superiores ou inferiores repetindo o padrão. Este padrão é conhecido como escala cromática (MÜLLER, 2015). Na Figura 1 ilustra-se um exemplo de duas diferentes grafias para representar uma escala cromática, utilizando a forma cifrada e a partitura musical.

Figura 1 – Duas diferentes grafias para a escala cromática.



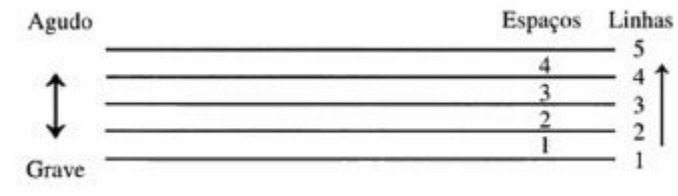
Fonte: Müller (2015)

2.3 Partitura musical

Uma partitura, ou pauta musical, constitui a reunião de cinco linhas horizontais paralelas e equidistantes, formando entre si quatro espaços, conhecida como penta-

grama, ilustrada na Figura 2. As linhas e os espaços são contatos de baixo para cima. As notas mais agudas são escritas na parte de cima da pauta, e as graves na parte de baixo (ALVES, 2004). A combinação da pauta com símbolos que representam as notas musicais forma a chamada partitura musical, na qual é possível transcrever as informações de uma música como velocidade, altura, ritmo, harmonia, entre outros atributos (PRIOLLI, 2015).

Figura 2 – Formação da pauta musical.



Fonte: Alves (2004)

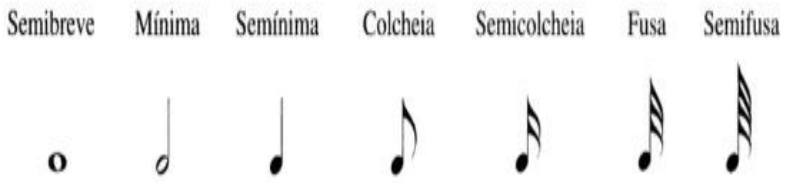
Os tempos ou momentos constituintes da pauta musical são agrupados em porções iguais, de dois em dois, de três em três ou de quatro em quatro. Os agrupamentos de quatro tempos são uma duplicação dos de dois tempos. Esses agrupamentos constituem unidades métricas chamadas de compasso. As barras verticais cortando linhas e espaços são conhecidas por barras de compasso ou travessão. As barras de compasso servem para distinguir a natureza do ritmo, ou seja, distinguir entre pulsos fortes e fracos. Para uma sucessão rítmica binária, o acento forte aparece a cada dois momentos, enquanto que para a ternária aparece a cada três momentos (POZZOLI, 1983). O compasso nada mais é do que o agrupamento ordenado de diversos momentos, em termos musicais representando a medida de tempo.

A acentuação métrica dos compassos de dois tempos se dá pela execução do primeiro tempo com acentuação forte e o segundo com acentuação mais fraca. Os compassos de três tempos se dão pela execução do primeiro tempo com acentuação forte, o segundo e o terceiro tempo com acentuação mais fraca. Os compassos de quatro tempos se dão pela execução do primeiro tempo com acentuação forte, o segundo com acentuação mais fraca, o terceiro com acentuação menos forte e o quarto com acentuação mais fraca (POZZOLI, 1983).

Uma figura que preenche um tempo é chamada de unidade de tempo, e a que

preenche um compasso é chamada de unidade de compasso. As unidades de tempo são suscetíveis de ser divididas em duas ou três partes iguais, daí a necessidade de distinguir a duração de tempo que ocupa todo o compasso (POZZOLI, 1983). O número $\frac{4}{4}$ no lado superior esquerdo na Figura 1 é uma fração que determina o número de tempos do compasso, e a figura que representa a unidade de tempo. Para determinar o nome da nota e sua altura na escala coloca-se, no princípio da pauta, um sinal chamado clave. Cada figura está associada a uma nota musical com seus respectivos valores de duração e altura. Na Figura 3 tem-se a representação de cada símbolo musical.

Figura 3 – Símbolos musicais.



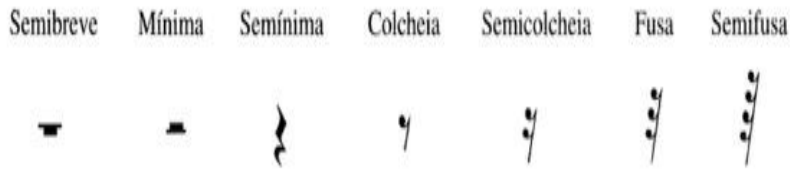
Fonte: Alves (2004)

Analisando a Figura 3, da esquerda para direita, temos os símbolos musicais que representam as durações: a semibreve, a mínima, a semínima, a colcheia, a semicolcheia, a fusa e a semifusa. A idéia de proporção presente entre elas vem a partir da duração em escala temporal. Uma semibreve representa uma unidade de tempo proporcional a duas mínimas, uma mínima representa o valor de tempo referente a duas semínimas, e assim por diante. Cada representação é referente à unidade de tempo padrão, ou seja, a semibreve. A semifusa, o último símbolo musical, equivale à menor proporção, representando $\frac{1}{64}$ do tempo de duração da semibreve (POZZOLI, 1983).

2.4 Pausas, ligaduras e ponto de aumento

As pausas são figuras que indicam a duração de silêncio entre os sons, com função rítmica que vem dar sentido na ausência de valor (PRIOLLI, 2015). Cada figura de som tem sua respectiva figura de pausa, como ilustrado na Figura 4.

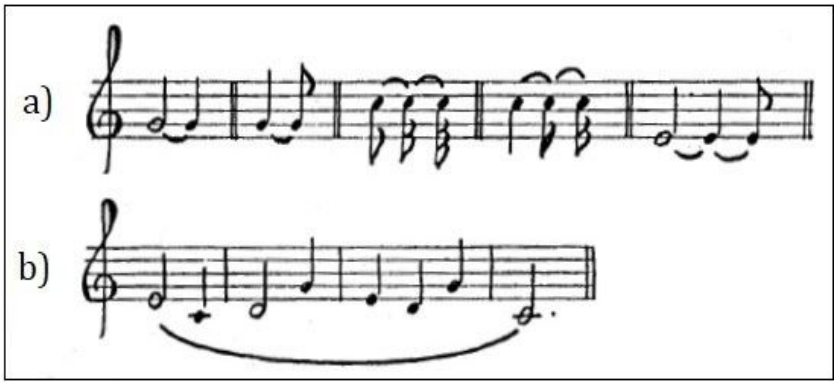
Figura 4 – Figuras de pausas.



Fonte: Alves (2004)

A ligadura é uma linha curva colocada sobre dois ou mais símbolos, indicando não haver interrupção do som. Quando colocada sobre sons de mesma entonação, o primeiro som deve ser emitido, sendo os demais uma prolongação do primeiro (PRIOLLI, 2015). Esta ligadura é conhecida como ligadura de valor. Em símbolos de entonação diferentes, a ligadura é conhecida como ligadura de portamento, não devendo haver uma interrupção na execução. Na Figura 5 tem-se um exemplo da ligadura de valor e outro da ligadura de portamento.

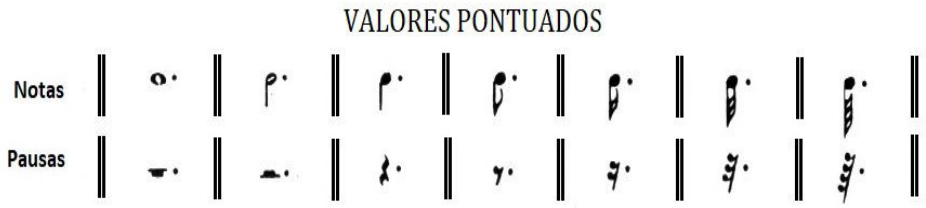
Figura 5 – Ligaduras: (a) de duração; (b) de portamento.



Fonte: Priolli (2015)

Um ponto colocado à direita de uma figura musical indica a execução com aumento da metade do valor dessa figura. Dois ou mais pontos podem ser colocados, onde cada ponto terá sempre a metade do valor de duração da nota ou ponto antecedente (PRIOLLI, 2015). A Figura 6 ilustra exemplos de figuras pontuadas.

Figura 6 – Figuras com ponto de aumento.



Fonte: Pozzoli (1983)

2.5 Conceitos básicos sobre som

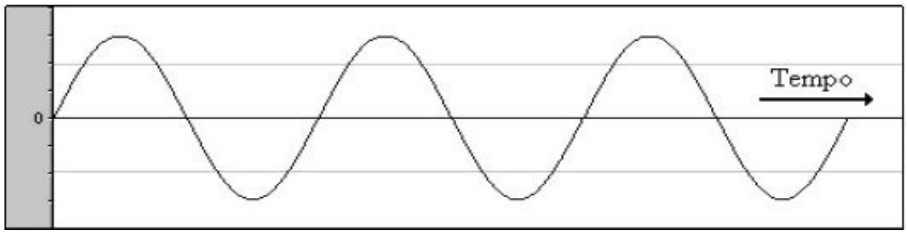
A relação entre física e música surgiu após a teoria ondulatória estabelecida nos séculos XVII e XVIII, sendo sedimentada mais tarde pelo matemático francês Jean-Baptiste Joseph Fourier quando desenvolveu no início do século XIX a formulação matemática utilizada na análise de qualquer fenômeno periódico (CARDOSO, 2010). A teoria ondulatória estabelece que cada perturbação física em um meio é um pulso, e uma sequência repetitiva e regular de pulsos constitui uma onda. A onda sonora é um exemplo de uma sequência regular de pulsos que se propaga pelo ar.

O som é um fenômeno vibratório resultante da pressão do ar, ou seja, mudanças na pressão que podem ser percebidas com nossos ouvidos. Cada mudança resulta em seções de ar de natureza mais densa, e outras rarefeitas, ocorrendo sucessivamente uma após outra e expandindo-se. Essas vibrações, quando mantêm um padrão repetitivo, podem ser interpretadas como formas de ondas periódicas (MILETTO et al., 2004).

Para a representação gráfica do som, as variações na pressão do ar podem ser interpretadas como formas de onda, conhecidas como senóides, mostrando as mudanças da pressão do ar conforme a passagem do tempo (MILETTO et al., 2004). Na Figura 7, ilustra-se uma representação gráfica do som com base em uma onda no chamado domínio temporal, onde a curva mais próxima da parte inferior do gráfico representa a pressão do ar mais baixa e a curva mais próxima da parte superior representa um aumento da pressão do ar.

A repetição de uma onda periódica é chamada de ciclo. O número de ciclos dentro de um intervalo de tempo em uma onda pode revelar características importantes

Figura 7 – Representação gráfica no domínio temporal.



Fonte: Miletto et al. (2004)

do som analisado (MILETTO et al., 2004). Existem diversos modelos funcionais de sinais para interpretação das características no domínio e no contra-domínio da função que representa o sinal. Geralmente, são utilizados dois modelos funcionais para essa representação: o modelo temporal e o modelo espectral. No modelo temporal, um sinal é determinado por uma função que define a variação do sinal no domínio do tempo. No modelo espectral, conhecido também como domínio frequencial, o sinal fica completamente caracterizado pela sua amplitude, frequência e pelo seu ângulo de fase. A frequência fornece uma medida da variação do sinal por unidade de tempo, ou seja , o sinal dá n ciclos completos por unidade de tempo (CARVALHO et al., 2009). Nas próximas seções são abordados os elementos básicos do som analisados por sua representação frequencial.

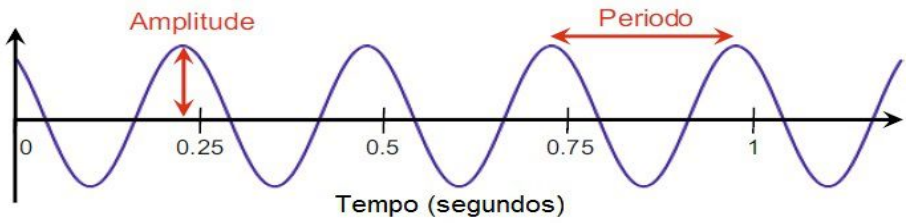
2.5.1 Representação frequencial do som

Em termos gerais, uma onda pode ser descrita como uma oscilação que viaja pelo espaço, onde a energia é transferida de um ponto para outro (MÜLLER, 2015). Essa oscilação, produzida por mudanças na pressão do ar, refere-se a uma repetição periódica de deformações e restaurações. O número de ciclos dentro de um intervalo de um segundo é chamado frequência e expresso em unidades chamadas *Hertz* (Hz) (MILETTO et al., 2004) em homenagem ao físico alemão Heinrich Hertz (1837-1894).

O período de uma onda é definido como o tempo para completar um ciclo. Uma senóide é especificada por sua frequência, amplitude e sua fase. A fase é determinada quando, dentro de um ciclo, a senóide está no tempo zero. Na Figura 8 ilustra-se um exemplo de uma onda senoidal com uma frequência de $4Hz$, na qual o período é

corresponde a um quarto de segundo (0,25s).

Figura 8 – Onda senoidal com frequência de $4Hz$.



Fonte: Müller (2015)

2.5.2 Altura, intensidade e timbre

Ao analisar uma representação de um som musical por meio de uma onda, percebe-se que sua frequência difere quando os sons mais agudos são comparados com os sons mais graves. Quanto mais agudo o som, maior o número de ciclos por unidade de tempo e, quanto mais grave, menor é este valor. Esta mudança, quando o som é agudo ou grave, medida em *Hertz*, é chamada de altura tonal (MILETTO et al., 2004). Uma faixa de som mais aguda é considerada como tendo maior altura, por conseguinte, uma faixa mais grave, como tendo menor altura.

Uma importante propriedade do som é sua intensidade, um termo geral que é usado para se referir ao seu volume (MÜLLER, 2015). As denominações “alto” e “baixo” devem ser utilizadas para distinguir a altura do som, evitando confundir com a intensidade ou volume. A mudança no volume do som pode ser identificada com a mudança de amplitude das ondas. Quanto maior a amplitude, mais forte é o som, quanto menor, o som é mais fraco (MILETTO et al., 2004).

Dois instrumentos musicais diferentes não produzem o mesmo som ao serem tocados com a mesma altura e a mesma intensidade. Isso é um aspecto fundamental chamado de timbre (MÜLLER, 2015). O timbre permite distinguir sons de mesma intensidade e altura, emitidos por fontes sonoras diferentes. Observando-se ondas de sons com timbres diferentes, nota-se que as formas das ondas diferem entre si (MILETTO et al., 2004). Na Figura 9 têm-se três tipos de ondas diferentes, seguidas de suas características e o instrumento que se assemelha a cada uma.

Figura 9 – Três tipos de onda, suas características de timbre, e instrumentos correspondentes.

Forma de onda	Timbre	Instrumento
Onda senoidal 	*suave, doce*	*flauta, assovio*
Onda dente-de-serra 	*claro, brilhante*	*violino, trompete*
Onda retangular 	*simples, "quente"*	*clarinete, oboé*

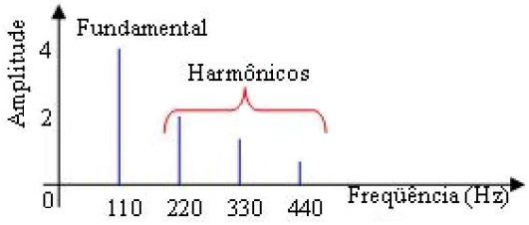
Fonte: Miletto et al. (2004)

2.5.3 Frequência fundamental

O som pode ser representado por uma soma de diversas ondas individuais, sendo que cada onda corresponde a uma determinada frequência múltipla da frequência inicial. O cálculo da frequência fundamental, conhecido como detecção do tom da nota, consiste em extrair de um sinal sonoro o componente de mais baixa frequência, isolando-a de outros componentes presentes. Os componentes múltiplos da fundamental são conhecidos como harmônicos, constituindo a chamada série harmônica (MILETTO et al., 2004).

A Figura 10 contém uma onda fundamental de $110Hz$ e sua série harmônica correspondente. É importante destacar que cada instrumento musical tem um número de harmônicos diferentes específicos, ao soar determinada frequência fundamental.

Figura 10 – Espectro de frequência com a fundamental de $110Hz$ e seus respectivos harmônicos.



Fonte: Ferreira (2006)

O tom de uma nota musical é determinado por sua frequência fundamental. A faixa de frequência entre dois tons é referida como intervalo. Chamam-se tons puros

os sons que não têm nenhum outro harmônico e consistem em uma só frequência simples. A forma de onda de um tom puro é caracterizada por uma onda senoidal, porém tons puros só podem ser criados artificialmente (MILETTO et al., 2004). Os primeiros harmônicos determinam o timbre do som, enquanto que os de ordem elevada caracterizam o "brilho". O conteúdo harmônico produzido por cada instrumento musical é, portanto, diferente, ainda que possua a mesma frequência fundamental (OLIVEIRA, 2007a).

A relação entre duas frequências, sendo uma mais alta do que a outra, com uma razão de $2:1$ é conhecida como oitava. As frequências de $220Hz$, $440Hz$, e $880Hz$, por exemplo, soam similares. Essa percepção de similaridade motivada pela notação de uma oitava se deve à natureza de percepção logarítmica natural pelo ouvido humano (MÜLLER, 2015). As frequências audíveis alcançadas pelos nossos ouvidos estão entre as faixas de $20Hz$ e $20.000Hz$.

Sabendo que a escala musical temperada tem doze intervalos por oitava, os intervalos de tons e semitons podem ser obtidos por uma determinada frequência f_0 multiplicada sucessivamente por um fator multiplicador, obtido pela equação 2.1, onde f_0 é frequência fundamental e k é índice da frequência a ser obtida.

$$f_k = f_0 \, 2^{\frac{k}{12}} \tag{2.1}$$

As frequências que compõem a oitava da escala musical são resultantes de uma progressão geométrica crescente, em escala logarítmica, sendo o primeiro termo a frequência inicial escolhida e a razão igual a $2^{\frac{1}{12}}$ (MÜLLER, 2015). Em outras palavras, dada uma frequência fundamental f_0, é possível obter as outras frequências (tons e semitons) da escala musical temperada. Sinais musicais são geralmente complexos, constituídos de múltiplos componentes frequenciais. Por causa dessa complexidade, a extração de características pode ser uma tarefa que requer o devido cuidado.

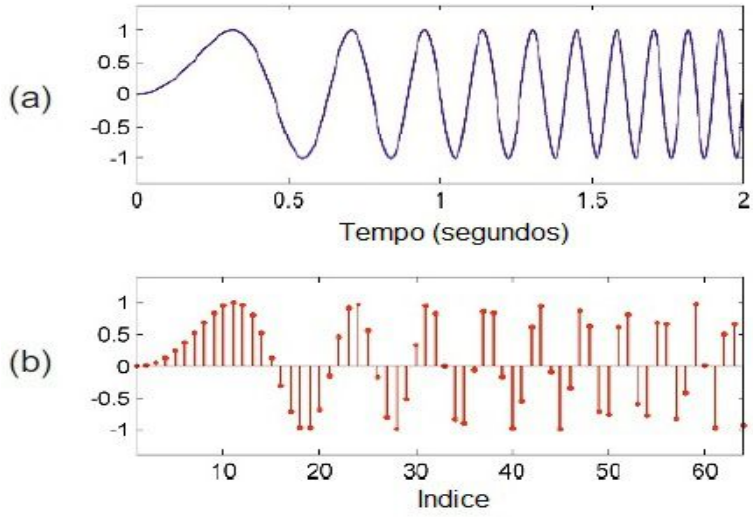
2.6 Sinais, amostragem e classificação

Um sinal é um conjunto de dados ou informação, dado por funções de uma variável independente de tempo ou de espaço. As operações de processamento de sinais podem ser abordadas de duas maneiras: a abordagem analógica e a abordagem

digital. Um sinal cuja amplitude pode assumir qualquer valor, em uma faixa contínua, em qualquer tempo, é um sinal analógico. Um sinal cuja amplitude pode assumir apenas um númeo finito de valores discretos é um sinal digital (LATHI, 2007).

Um sinal em tempo contínuo $x(t)$ pode ser convertido para um sinal de tempo discreto $x[n]$ a partir da técnica de amostragem. A taxa de amostragem f_s deve ser mantida suficientemente alta para permitir a reconstrução do sinal com o mínimo de erros (LATHI, 2007). O processo de amostragem é realizado pela representação de N pontos de amplitude do sinal $x(t)$ por números proporcionais a cada amplitude. Um som pode então ser representado por uma sequência de números em que cada um representa uma amplitude em um único instante de tempo (MILETTO et al., 2004). Na Figura 11 tem-se um sinal contínuo no tempo e sua representação amostrada. Uma

Figura 11 – Amostragem de um sinal. a) Um sinal $x(t)$; b) sua representação amostrada usando frequência amostral de $32Hz$.



Fonte: Müller (2015)

função de equidistância é aplicada para conversão do sinal contínuo para sinal discreto, dado pela equação 2.2 na qual $x(n)$ é uma amostra no tempo n e T é o período de amostragem.

$$x(n) = f(n.T) \tag{2.2}$$

A frequência amostral f_s dada por $f_s = \frac{1}{T}$ é chamada de taxa amostral. Para a amostragem realizada por um computador digital, um processo de quantização é necessário quando o nível de uma determinada amostra deve permanecer entre dois

valores digitais. A quantização consiste em aproximar tal nível do valor digital mais próximo existente (MILETTO et al., 2004). O número de bits usados para representar cada amostra coletada refere-se à resolução. Cada bit acrescentado na resolução dobra o número de valores usados para representar a variação de amplitude da onda.

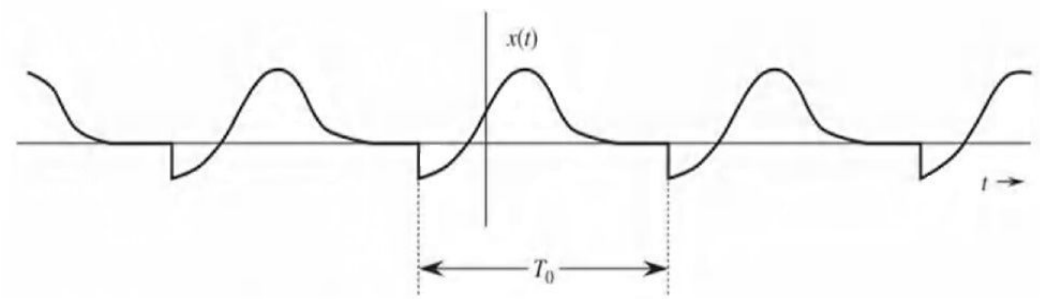
O principal problema na técnica de amostragem está relacionado com o número de amostras por segundo que devem ser colhidas. A análise deste problema é feita no Teorema de Shannon-Nyquist (OLIVEIRA, 2007a). O teorema estabelece que um sinal com banda limitada em f_sHz pode ser univocamente representado pelas suas amostras discretas, se estas forem colhidas em uma taxa de pelo menos $2f_s$ amostras equiespaçadas por segundo. Isto significa que um sinal original analógico pode ser reconstruído se o valor da resolução amostral f_s for o dobro da maior frequência encontrada no sinal (MÜLLER, 2015). Esse limite de banda do sinal é conhecido como limite ou frequência de Nyquist. Ao se tentar reproduzir uma frequência maior do que a frequência de Nyquist, ocorre um fenômeno chamado *aliasing*. A frequência é "espelhada" ou "rebatida" para uma região mais grave do espectro.

Para aumentar a resolução do sinal analisado, é necessário aumentar o número de pontos em um determinado espaço de tempo, permitindo assim um melhor detalhamento na representação do sinal. Porém, qualquer esquema de análise em tempo-frequência está sujeito à sua resolução amostral: um sinal não pode ser analisado com ambas as resoluções, alta no tempo e alta na frequência. O aumento da resolução de um implica a diminuição da resolução de outro e vice-versa (JUILLERAT; ARISONA; SCHUBIGER-BANZ, 2008). Esse problema é conhecido como Princípio da Incerteza de Heisenberg, formulado em 1946 por Gabor-Heisenberg. Trata-se de uma relação entre a duração efetiva de um sinal e sua banda passante efetiva, obtida no contexto de sinais determinísticos. Heisenberg concluiu que é impossível estabelecer uma determinada frequência exata e o tempo exato da ocorrência dessa frequência no sinal (OLIVEIRA, 2007b).

2.6.1 Classificação de sinais

A classificação de sinais pode ser baseada na maneira como eles são definidos em função do tempo. Um sinal $x(t)$ é um sinal de tempo contínuo se ele for definido

Figura 12 – Um sinal periódico com período T_0



Fonte: Lathi (2007)

para todo tempo t, cuja amplitude varia continuamente no tempo. Um sinal no tempo discreto é definido somente em instantes isolados, frequentemente derivado de um sinal de tempo contínuo por meio da técnica de amostragem (HAYKIN; VEEN, 2001). Um sinal periódico $x(t)$ é uma função que satisfaz a seguinte condição:

$$x(t) = x(t+T), \qquad \forall t \tag{2.3}$$

em que T é alguma constante positiva. O menor valor de T satisfaz a condição de periodicidade da equação 2.3, definindo a duração de um ciclo completo de $x(t)$, chamado de período fundamental (HAYKIN; VEEN, 2001). O período fundamental T descreve quão frequentemente um sinal periódico $x(t)$ se repete. A Figura 12 contém um sinal periódico $x(t)$ com período T_0. Um sinal $x(t)$ é chamado aperiódico ou não-periódico quando não existe nenhum valor de T para satisfazer a condição da equação 2.3.

3 REVISÃO BIBLIOGRÁFICA - PARTE II: ANÁLISE TEMPO-FREQUÊNCIA DE SINAIS E REDES NEURAIS ARTIFICIAIS

Este Capítulo descreve um breve resumo das principais técnicas utilizadas na área de análise de sinais, focalizando em particular as técnicas adequadas à análise de sinais sonoros. São apresentadas a Taxa de Cruzamento por Zero, a Autocorrelação, seguidas dos conceitos da Transformada Discreta de Fourier e a Transformada Discreta de Fourier de Tempo Curto, e a Transformada Q-Constante. Aborda-se uma breve introdução à teoria Wavelet na representação de sinais, a Transformada de Wavelet Discreta e a Transformada Wavelet-Packet utilizada nesse estudo. Por fim, são apresentados os conceitos básicos de Redes Neurais Artificiais (RNAs) que posteriormente serão aplicados no desenvolvimento deste trabalho.

3.1 Taxa de cruzamento por zero

A Taxa de Cruzamento por Zero (ZCR - *Zero Crossing Rate*) é um método de detecção da frequência fundamental que compõe um sinal (CHEN; SHEN; HSU, 2015). Em um sinal periódico, o ZCR é a taxa de mudanças que ocorrem na amplitude do sinal, ou seja, quando este corta o eixo zero passando de um valor positivo para um valor negativo e vice-versa, dividido pelo tamanho N do quadro (um segmento do sinal analisado). Na figura 13 tem-se, como exemplo, um segmento de sinal com nove cruzamentos no eixo zero.

Figura 13 – Um sinal com nove cruzamentos no eixo zero.

Fonte: McLoughlin (2009)

O ZCR pode ser definido de acordo com a seguinte equação (MCLOUGHLIN, 2009):

$$Z(i) = \frac{1}{2W} \sum_{n=1}^{W} |sgn[x_i(n)] - sgn[x_i(n-1)]|, \tag{3.1}$$

onde sgn é uma função de sinal (GIANNAKOPOULOS; PIKRAKIS, 2014), isto é

$$sgn[x_i(n)] = \begin{cases} 1, x_i(n) \geq 0, \\ -1, x_i(n) < 0. \end{cases} \tag{3.2}$$

Assumindo que um segmento do sinal pode ser representado por um vetor contendo amostras do sinal sob análise, ao percorrer o vetor contabiliza-se o número de cruzamentos que é finalmente dividido pelo tamanho do segmento, retornando a taxa de cruzamentos.

O ZCR é útil em numerosas aplicações, incluindo detecção de voz, gêneros musicais, quadros silenciosos entre outros (SHETE; PATIL; PATIL, 2014). Um exemplo simples de utilização do ZCR é um algoritmo de detecção de voz no qual um baixo valor de ZCR indica um sinal contendo voz enquanto que um valor mais alto pode indicar um sinal ruidoso (GIANNAKOPOULOS; PIKRAKIS, 2014).

Algoritmos baseados em ZCR são simples e rápidos, porém são muito restritivos, uma vez que um sinal ruidoso pode causar um distúrbio na determinação dos pontos de cruzamento por zero (CHEN; SHEN; HSU, 2015). Um sinal sonoro geralmente apresenta um certo nível de ruído de fundo, resultando em altas taxas de cruzamento por zero, mesmo em sinais silenciosos. Por essa razão, pode ser necessário incluir um limiar de tolerância na função de ZCR para tentar minimizar este problema (MEDHI; TALKUDHAR, 2014).

3.2 Autocorrelação

A autocorrelação possibilita quantificar o grau de semelhança entre o sinal e ele mesmo deslocado no tempo (TREVILLATO; BARBEDO; LOPES, 2005). Quando o sinal apresenta uma periodicidade, o deslocamento apresentará maior semelhança equivalente ao período da onda. As localizações dos picos da função de autocorrelação indicam periodicidades (conteúdo rítmico) mais destacadas na envoltória do sinal (SHIMAMURA; KOBAYASHI, 2001). A autocorrelação pode ser obtida pela seguinte equação:

$$AC = \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n+\tau) \tag{3.3}$$

onde $x(n)$ é o sinal de entrada e τ é o fator de deslocamento. Utilizando um processo de deslocamento de janelas, é possível localizar a frequência fundamental f_0 do sinal. Para cada janela aplica-se a autocorrelação e calcula-se a fundamental, utilizando a equação 3.4:

$$f_0 = f_s/n_d \tag{3.4}$$

onde f_s é a frequência de amostragem e n_d é o deslocamento (índice) do primeiro pico da autocorrelação (TREVILLATO; BARBEDO; LOPES, 2005).

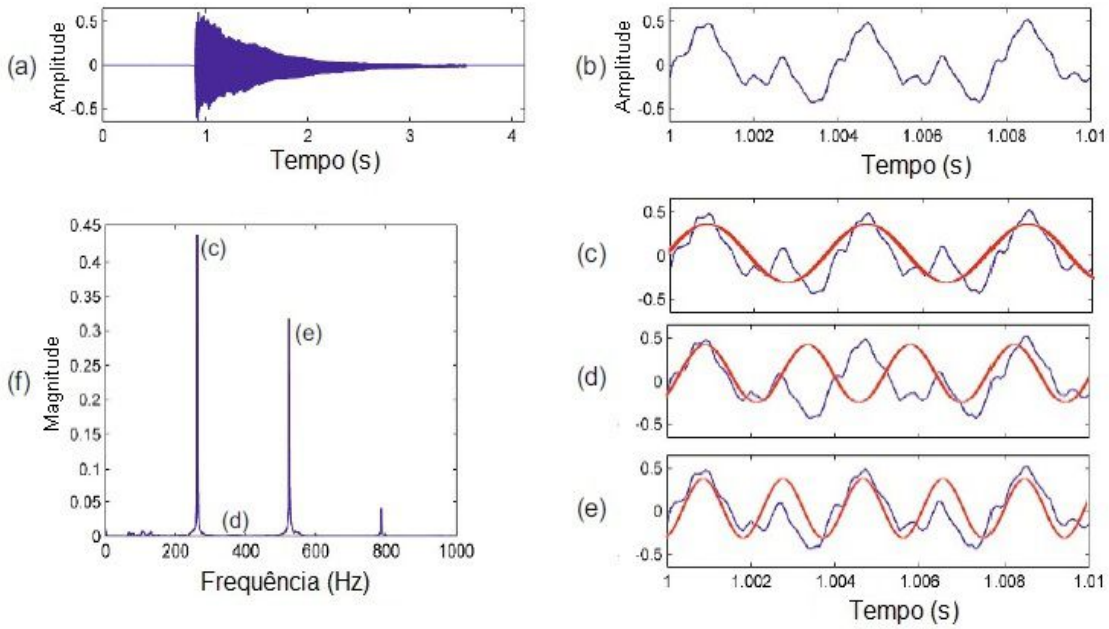
Dependendo da fonte sonora, o comportamento da autocorrelação pode apresentar grande variação. Problemas na análise podem surgir em razão da presença de harmônicas cuja energia relativa é grande comparada à frequência fundamental. Nesse caso outros picos podem surgir na autocorrelação antes daquele que corresponde a f_0. Uma alternativa para tentar minimizar o problema é a aplicação de limiares para reduzir o número de picos que não correspondem à fundamental. Porém, dependendo da quantidade de conteúdo harmônico encontrado, esse método nem sempre é eficiente para sinais mais complexos.

3.3 Transformada de Fourier de Tempo Discreto

Para a representação de um sinal, pode-se, em geral, utilizar dois domínios distintos: temporal e espectral. Na Figura 14 tem-se um sinal de áudio que representa uma única nota tocada por um piano. Para determinar o conteúdo frequencial, deve-se lembrar que o tom de uma nota musical está relacionado com sua frequência fundamental, sendo necessário determinar as principais oscilações periódicas do sinal (MÜLLER, 2015). Considerando um segmento do sinal de apenas $10ms$, com base na figura 14(b), pode-se observar que o sinal se comporta de maneira quase periódica.

As Figuras 14(c), 14(d) e 14(e) representam comparações com senóides de várias frequências. Sobrepondo o sinal da Figura 14(b) com uma senóide de $300Hz$, pode-se observar por meio da Figura 14(c), as três principais oscilações, representando três ciclos de oscilação dentro da seção de $10ms$, significando que o sinal contém um componente frequencial de aproximadamente $300Hz$. Comparando um sinal com

Figura 14 – Um sinal: (a) de uma nota tocada por um piano; (b) segmento do sinal com $10ms$; (c-e) Comparação do sinal com senóides de várias frequências; (f) Coeficientes de magnitude de cada comparação.



Fonte: Müller (2015)

senóides de várias frequências ω_s, obtém-se um coeficiente de magnitude d_{ω_s} relativo a cada senóide. Coeficientes com valores consideravelmente altos representam uma grande semelhança entre o sinal e senóide de frequência ω_s. Por outro lado, coeficientes com valores menores representam que o sinal não contém um componente periódico da frequência (MÜLLER, 2015). O espectro frequencial deste exemplo pode ser visto na Figura 14(f).

O estudo de sinais e sistemas usando representações senoidais é denominado Análise de Fourier (HAYKIN; VEEN, 2001). Com base nos coeficientes de Fourier, pode-se obter informações referentes a cada componente frequencial de um determinado sinal. Para cada classe diferente de sinais, existem quatro representações distintas de Fourier aplicadas, onde cada classe é definida por suas propriedades de periodicidade do sinal. Um sinal em tempo contínuo, quando periódico, pode ser representado pela Série de Fourier (FS - *Fourier Series*), já um sinal não periódico em tempo contínuo pode ser representado pela Transformada de Fourier (FT - *Fourier Transform*). A Série de Fourier de Tempo Discreto (DTFS - *Discrete Time Fourier Series*) aplica-se a sinais periódicos de tempo discreto e, finalmente, a Transformada de Fourier de Tempo

Discreto (DTFT -*Discrete Time Fourier Transform*) se aplica a sinais não periódicos de tempo discreto (HAYKIN; VEEN, 2001).

Cálculos da Transformada de Fourier necessitam de valores amostrados, pois um computador digital pode trabalhar somente com dados discretos. Quando um sinal $x(t)$ é amostrado, precisa-se relacionar as amostras do sinal $x(t)$ no domínio temporal com as amostras de $X(\omega)$ no domínio frequencial (LATHI, 2007). Em outras palavras, é necessário um mapeamento do sinal $x(t)$ que depende de uma variável discreta de tempo t numa transformada que depende de uma variável também discreta de frequência ω (DINIZ; SILVA; NETTO, 2014). Esse mapeamento é conhecido como a Transformada de Fourier de Tempo Discreto (DTFT).

A DTFT é a principal representação usada para aplicações computacionais, podendo ser aplicada para analisar combinações de sinais não periódicos de tempo discreto (HAYKIN; VEEN, 2001). Na prática, os dados são disponibilizados sob a forma de uma função do tempo de amostragem, representada por uma série temporal das amplitudes, separados por intervalos de tempo fixos, de duração limitada, transformando um sinal do domínio do tempo para o domínio da frequência.

Uma sequência $x(t)$ de um sinal contínuo no tempo pode ser convertida para o domínio espectral pela Transformada de Fourier (FT), conforme a equação 3.5:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}, \tag{3.5}$$

na qual percebe que tal caracterização no domínio frequencial depende da variavel contínua ω (DINIZ; SILVA; NETTO, 2014). Amostrando uniformemente a variável frequencial contínua ω, pode-se relacionar as amostras do sinal $x(t)$ no domínio do tempo com as amostras de $X(\omega)$ no domínio da frequência, preservando as informações da equação 3.5. Como $X(e^{j\omega})$ é periódica com período 2π, o processo de amostragem é realizado tomando-se N amostras espaçadas linearmente, entre 0 e 2π. Utilizando as frequências $\omega_k = (\frac{2\pi}{N})k$, para $k = 0, 1, ..., N-1$, onde $x(t)$ tem duração finita, e suas amostras não nulas estão dentro do intervalo $0 \leq t \leq N$, a DTFT pode ser definida pela equação 3.6:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}kn} \tag{3.6}$$

para $k \in [0 : N-1]$, onde $X = (X(0), X(1), ...X(N-1))$ denotam o vetor de coeficientes complexos de Fourier (MÜLLER, 2015). A exponencial pode ser substituída por

$$\left[\cos\left(\frac{2\pi}{N}kn\right) - j.\sin\left(\frac{2\pi}{N}kn\right)\right]. \tag{3.7}$$

Esta substituição na exponencial, usada na equação, é conhecida como fórmula de Euler (MÜLLER, 2015). Fazendo $X(k) = a_k + jb_k$ pode-se escrever

$$\sum_{n=0}^{N-1} x(n).\cos\left(\frac{2\pi}{N}kn\right) - j.\sum_{n=0}^{N-1} x(n).\sin\left(\frac{2\pi}{N}kn\right) \tag{3.8}$$

e então

$$a_k = \sum_{n=0}^{N-1} x(n).\cos\left(\frac{2\pi}{N}kn\right) \quad b_k = \sum_{n=0}^{N-1} x(n).\sin\left(\frac{2\pi}{N}kn\right) \tag{3.9}$$

onde a_k e b_k representam respectivamente valores reais e valores imaginários do coeficiente $X(k)$. Pode-se então obter a amplitude e o ângulo de fase respectivamente das relações:

$$A = \sqrt{a_k^2 + b_k^2} \tag{3.10}$$

$$\phi = arctg\left(\frac{b_k}{a_k}\right) \tag{3.11}$$

sendo A o valor da magnitude no domínio frequencial e ϕ o ângulo de fase (ROZINAJ; NAGY, 2004). Computando a_k e b_k , pode-se designar k para a relação que determina a frequência real, dado em

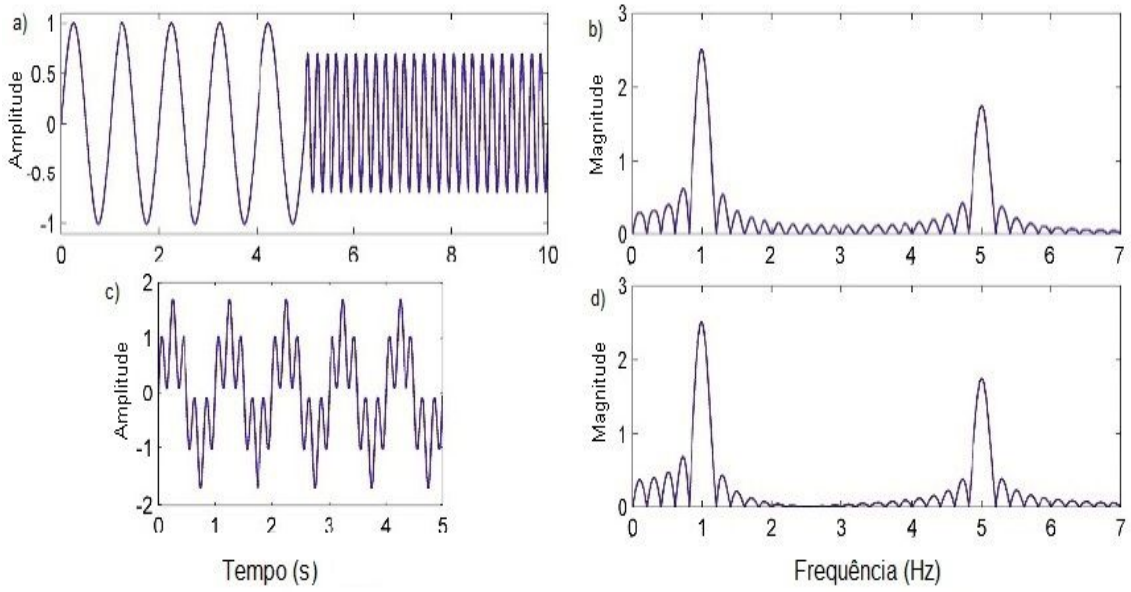
$$k = \frac{f}{\frac{f_s}{N}} = \frac{f.N}{f_s}, \tag{3.12}$$

onde N é o comprimento da DTFT, f é frequência detectada e f_s é a frequência amostral.

Uma limitação da DTFT é o considerável número de operações aritméticas envolvidas no seu cálculo (DINIZ; SILVA; NETTO, 2014). Em 1965, Cooley e Tukey propruseram algoritmos mais eficientes para os cálculos da DTFT. A Transformada Rápida de Fourier (FFT - *Fast Fourier Transform*) é uma eficaz implementação computacional da DTFT, possibilitando obter uma estimativa da amplitude fundamental e as suas harmônicas com uma aproximação razoável. A FFT utiliza algoritmos inteligentes para o cálculo da DTFT em um tempo computacional mais baixo (INGALE, 2014).

Na Figura 15(a) tem-se um sinal com duas senóides subsequentes e o seu respectivo espectro com frequências de $1Hz$ e $5Hz$ em 15(b). A sobreposição das mesmas senóides e seu espectro pode ser visto respectivamente nas figuras 15(c) e 15(d). Nota-se que os dois sinais são diferentes em natureza, mas os coeficientes de magnitude são mais ou menos os mesmos. Isso demonstra a limitação da Transformada de Fourier ao analisar sinais com características mutáveis ao longo do tempo (MÜLLER, 2015).

Figura 15 – Dois diferentes sinais no domínio do tempo e seus respectivos espectros de Fourier. (a) Dois sinais subsequentes de frequência $1Hz$ e $5Hz$; (b) Espectro de Fourier de (a); (c) Sobreposição dos dois sinais; (d) Espectro de Fourier de (c).



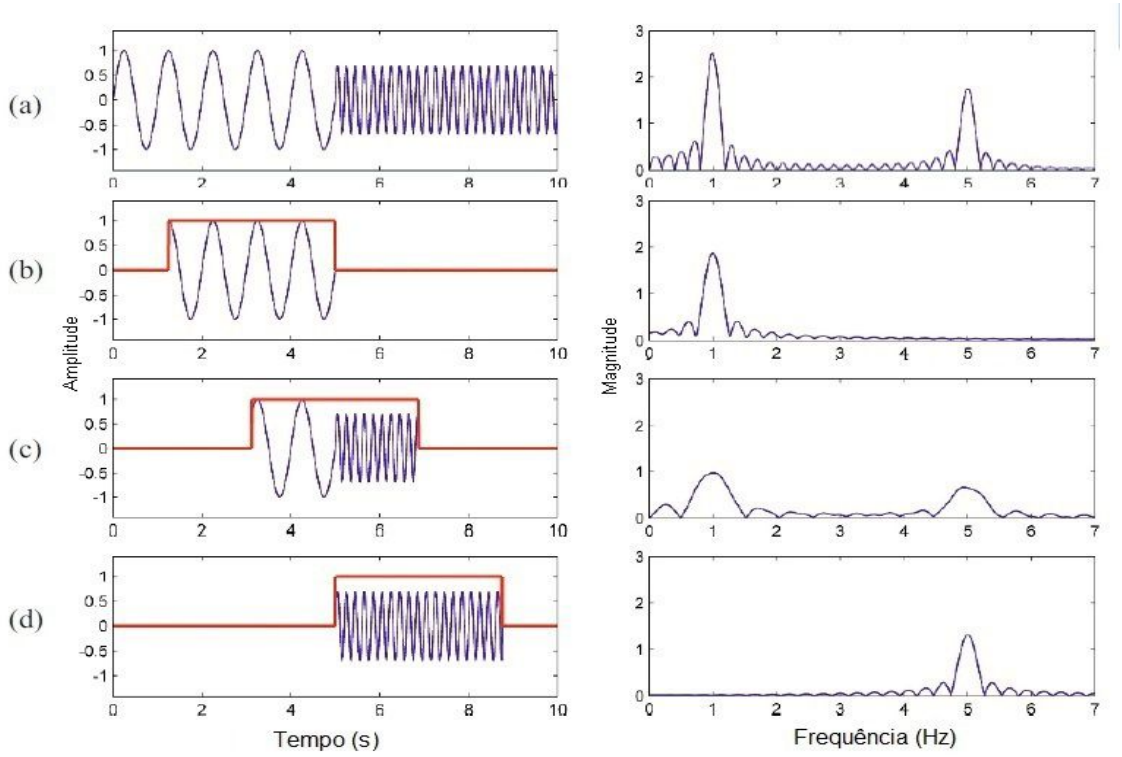
Fonte: Müller (2015)

3.4 Transformada de Fourier Janelada

A Transformada de Fourier analisa um sinal levando em consideração a duração completa, de $-\infty$ a $+\infty$. Entretanto, para certas aplicações, precisamos calcular o conteúdo frequencial local do sinal em certa região (DINIZ; SILVA; NETTO, 2014). Para recuperar a informação do conteúdo frequencial local, Dennis Gabor introduziu no ano de 1946, a Transformada de Fourier Janelada (STFT - *Short-Time Fourier Transform*). Em vez de considerar um sinal como um todo, a ideia principal do STFT é considerar apenas uma pequena parte do sinal (MÜLLER, 2015).

A STFT baseia-se na DTFT, tratando uma parte de um sinal não-estacionário como um sinal estacionário, analisando o sinal por meio de uma janela temporal fixa, estreita o suficiente para que a parte do sinal visto a partir da janela seja estacionário. A Figura 16(a) contém uma ilustração que se refere á obtenção local do sinal original e seu respectivo espectro, multiplicando o sinal com funções de janela retangular adequadamente deslocada; na Figura 16(b) a seção local resultante na janela contém uma frequência de $1Hz$ e pico espectral de $\omega = 1$; em 16(c) a janela deslocada novamente contém uma frequência de $1Hz$ e componentes de $5Hz$, refletidos pelos dois picos $\omega = 1$ e $\omega = 5$; e finalmente a janela deslocada na Figura 16(d) contém apenas o conteúdo frequencial de $5Hz$. A equação da STFT usando uma expressão

Figura 16 – Um sinal no domínio do tempo e seu respectivo espectro no domínio da frequência.



Fonte: Müller (2015)

tempo-frequência pode ser definida como:

$$X(m,k) = \sum_{n=0}^{N-1} x(n+mH)w(n)e^{-j\frac{2\pi}{N}kn} \tag{3.13}$$

onde o número complexo $X(m,k)$ denota a STFT do sinal analisado para a m-ézima janela e o k-ézimo coeficiente Fourier, $x(n)$ é a função de janelamento, N o número

de amostras em uma janela e H determina o tamanho do passo no qual a janela é deslocada ao longo do sinal (MÜLLER, 2015).

Para cada janela m fixada, um vetor espectral de tamanho $k+1$ é dado pelos coeficientes $X(m,k)$ calculados pela DTFT. Para a dimensão temporal, cada coeficiente é associado com uma posição física no tempo:

$$T_{coef(m)} = \frac{m.H}{f_s} \tag{3.14}$$

dada em segundos (MÜLLER, 2015). Para a dimensão frequencial, o coeficiente k corresponde à frequência

$$F_{coef(k)} = \frac{k.f_s}{N} \tag{3.15}$$

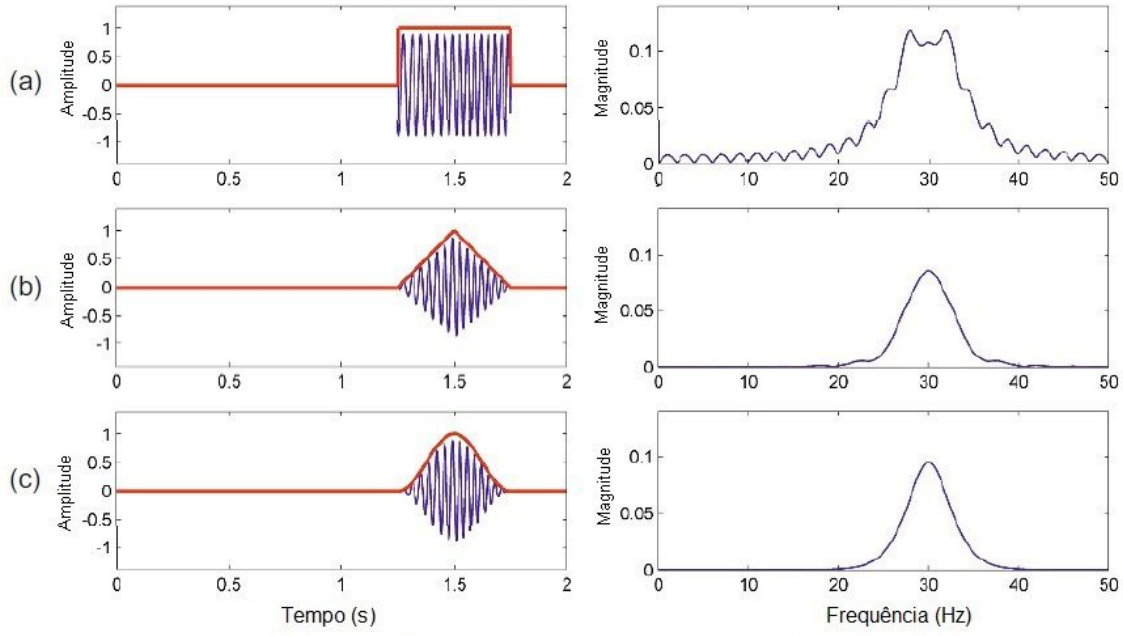
dada em Hertz (Hz) (MÜLLER, 2015). O espectrograma, denotado por Y, representando duas dimensões, temporal e frequencial, é dado pela equação 3.16.

$$Y(m,k) = |X(m,k)|^2 \tag{3.16}$$

Analisando um sinal musical como exemplo, dada uma janela de tamanho $N = 4096$ e um passo temporal $H = \frac{N}{2}$, com uma frequência amostral $f_s = 44100Hz$, com base na equação 3.14, a resolução temporal é $\frac{H}{f_s} \approx 46.4ms$ e a resolução frequencial é $\frac{f_s}{N} \approx 10.8Hz$, conforme a equação 3.15. Para obter uma melhor resolução frequencial, bastaria aumentar o tamanho N da janela, porém isso leva a uma resolução temporal mais baixa, perdendo-se a capacidade de capturar fenômenos locais no sinal (MÜLLER, 2015).

Uma STFT é função tanto do sinal de entrada quanto da função de janela. A função de janelamento afeta diretamente a estimativa espectral do sinal calculado pela STFT. O projeto para a escolha da janela é uma ciência por si só. O uso de janelas retangulares é mais simples, porém traz alguns incovenientes como descontinuidades nos limites, que, ao invés de fazer parte do sinal original, estes componentes frequenciais vêm a partir das propriedades da janela retangular (MÜLLER, 2015). Para atenuar esses efeitos, outras janelas podem ser usadas, devendo ser feita uma avaliação prévia para a escolha que melhor se adapte à análise. Na Figura 17 tem-se um sinal no tempo e o espectro de magnitude de Fourier usando diferentes funções de janelamento.

Figura 17 – Um sinal no domínio do tempo e a representação espectral usando diferentes funções de janelamento. (a) janela retangular; (b) janela triangular e (c) janela de Hann.



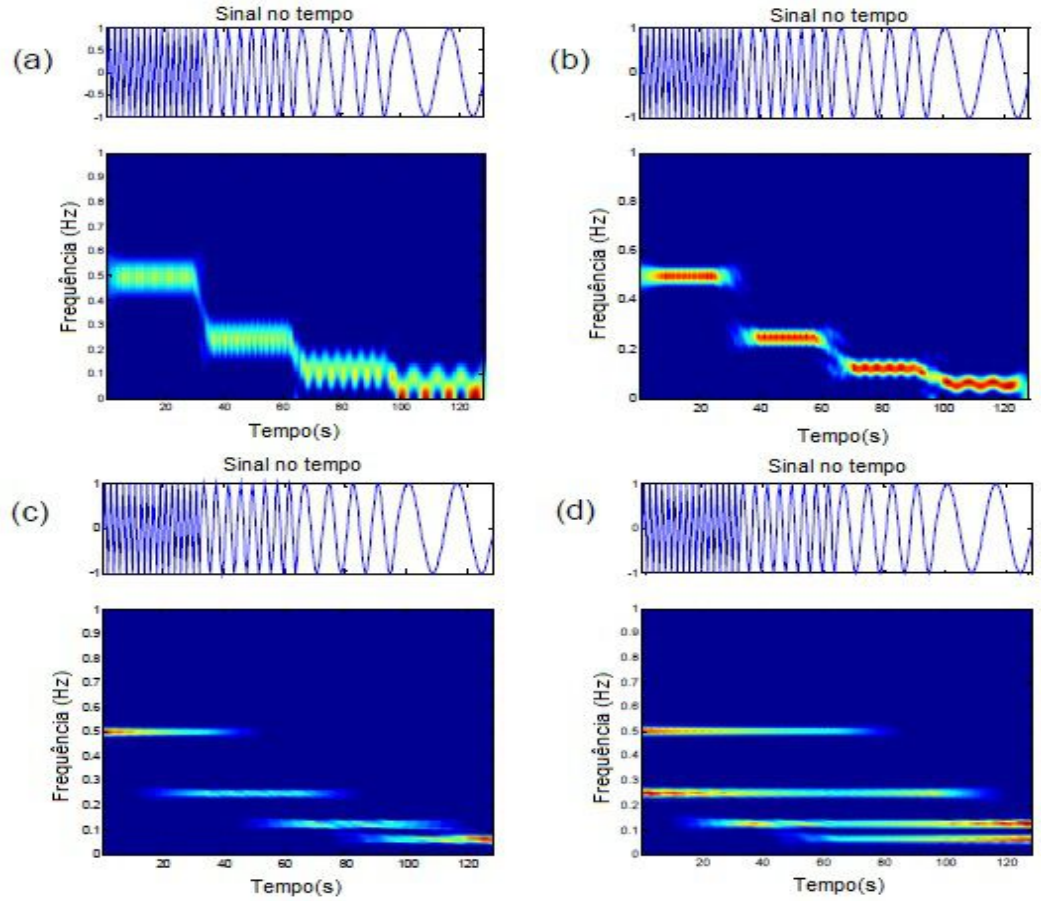
Fonte: Müller (2015)

Em sinais altamente não estacionários, as resoluções temporal e frequencial em janelas fixas da STFT podem se tornar uma grave desvantagem. O problema é a escolha da largura para a função de janelamento. Isso porque eventos de durações muito diferentes requerem graus variados de resolução (DINIZ; SILVA; NETTO, 2014). Por exemplo, suponha que sejam usadas janelas retangulares cada uma com um número de M de amostras. Cada janela dá origem a M coeficientes Fourier e, sendo assim, a preservação do número total de amostras N requer o uso de $\frac{N}{M}$ janelas. Quando o número de coeficientes M aumenta, melhora-se a análise de cada janela (maior número de coeficientes Fourier dentro de uma janela), mas perde-se resolução espacial (menor número de janelas). Esse problema é conhecido como Princípio da Incerteza de Heinsenberg (GALVÃO et al., 2001) (ver seção 2.6). Uma boa resolução temporal pode funcionar bem para sons transitórios, enquanto uma boa resolução frequencial pode funcionar bem para sons constantes, mas nenhuma escolha oferece bons resultados em sinais com ambos os sons transitórios e constantes (JUILLERAT; ARISONA; SCHUBIGER-BANZ, 2008).

A Figura 18 contém um sinal concatenado por quatro diferentes frequências,

cada uma composta por 64 amostras: $0,5Hz$, $0,250Hz$, $0,125Hz$ e $0,0625Hz$. Nos espectrogramas das Figuras 18(a) e 18(b) foram utilizadas janelas com 15 e 31 amostras respectivamente. É possível observar que, utilizando uma janela de análise curta no tempo, consegue-se uma boa resolução no tempo, no entanto, sem obter uma boa resolução em frequência. As Figuras 18(c) e 18(d) utilizam janelas mais longas, de 65 e 127 amostras respectivamente. O resultado obtido possui uma rica resolução em frequência, mas pobre resolução no tempo.

Figura 18 – Espectrograma de um sinal composto por quatro senóides concatenadas de frequências de $0.5Hz$, $0.250Hz$, $0.125Hz$ e $0.0625Hz$ respectivamente. O tamanho das janelas: em (a) 15 amostras; (b) 31 amostras; (c) 63 amostras e (d) 127 amostras.



Fonte: Auger et al. (1996)

3.5 Transformada Q-Constante

Sinais discretos podem ser representados no domínio da frequência por meio da DFTS, ou com base na STFT quando se deseja uma análise multirresolução em

tempo-frequência. Porém, os resultados são dispostos sobre uma escala linear de frequências, analisando baixas e altas frequências em uma mesma resolução.

A Transformada Q-Constante (CQT - *Constant-Q Transform*) proposta por Brow refere-se a uma técnica que transforma um sinal no domínio do tempo $x(t)$ para o domínio tempo-frequência, em que as frequências centrais são geometricamente espaçadas, assim como nas escalas temperadas da música ocidental (BROWN; PUCKETTE, 1992). Os resultados são dispostos em uma escala logarítmica, diferentemente da DFTS em que as frequências são dispostas em uma escala linear. Isso permite definir uma análise tempo-frequência com frequências centrais arbitrárias e resolução de frequência arbitrária. Uma razão (o Q-fator) entre as frequências adjacentes deve ser escolhida de acordo com a precisão desejada. No caso da análise de sinais musicais, a razão escolhida deve ser a mesma na qual as frequências fundamentais das notas na escala estão dispostas (SZCZUPAK; BISCAINHO; CALÔBA, 2006).

A CQT pode ser vista como uma STFT, mas com frequências espaçadas logaritmicamente, por meio da variação do comprimento da janela de análise (ELOWSSON; FRIBERG, 2013), permitindo analisar sinais musicais com uma resolução frequencial alta o suficiente para separar os diferentes tons dentro de cada oitava na escala musical (FILLON; PRADO, 2012).

Dada uma frequência mínima f_0 de um sinal, o centro frequencial para cada banda pode ser obtido em:

$$f_k = f_0 2^{\frac{k}{b}} \ (k = 0, 1, ...) \tag{3.17}$$

onde o parâmetro b é o número de componentes, e k é o k-ésimo componente a ser calculado (CRANITCH; CYCHOWSKI; FITZGERALD, 2006). A proporção fixa de frequência central relativa a largura de banda é dada por:

$$Q = (2^{\frac{1}{b}-1})^{-1}. \tag{3.18}$$

Para aplicações musicais, o cálculo de f_k pode ser baseado nas frequências da escala de igual temperamento:

$$q = 2^{\frac{1}{12\beta}}, \qquad (\beta = 1, 2, 3, ...) \tag{3.19}$$

onde o fator β define a resolução espectral em frações de semitom (BROWN; PUCKETTE, 1992). Um comprimento de janela N_k, a uma frequência amostral f_s , pode ser obtido por

$$N_k = Q\frac{f_s}{f_k} \tag{3.20}$$

onde f_k é a frequência sob análise, obtendo uma largura de banda desejada para cada banda de frequência. Assim, a CQT é definida pela equação 3.21:

$$X(k) = \frac{1}{N_k}\sum_{n=0}^{N_k-1} x(n)W_{N_k}(n)e^{-j\frac{2\pi}{N_k}Qn} \tag{3.21}$$

onde $x(n)$ é um sinal no domínio temporal e W_{N_k} é uma função de janela de tamanho N_k (CRANITCH; CYCHOWSKI; FITZGERALD, 2006). Para obter seletividade constante e espaçamento logarítmico, o comprimento de janela (número de amostras analisadas) deve variar de acordo com a frequência desejada. O índice frequencial Q presente na exponencial fornece essa seletividade (SZCZUPAK; BISCAINHO; CALÔBA, 2006).

Uma desvantagem da CQT é o fato de que, para produzir uma resolução de frequência que seja adequada para a análise musical, uma variação no tamanho da janela de análise é necessária. Para calcular a CQT, o número de amostras utilizadas em uma janela de comprimento N_k depende do valor de cada componente calculado na equação 3.21, ou seja, quanto menor o valor de f_k, maior é valor de N_k. Isso significa que quanto menor o valor da componente frequencial, maior o comprimento de janela e vice-versa. Além disso, o número de amostras sobrepostas depende tanto da quantidade de amostras de um passo temporal H, quanto da componente f_k analisada (SZCZUPAK; BISCAINHO, 2009).

Um passo temporal H com um comprimento máximo igual ao comprimento do menor valor de N_k é necessário para que todas as amostras de sinal sejam analisadas para cada componente. Entretanto, isso gera um elevado custo computacional para os cálculos decorrentes do passo H reduzido, e haverá grande sobreposição entre intervalos de janelas consecutivas para frequências mais baixas, resultando em uma análise redundante. Uma alternativa seria optar por um passo H de comprimento intermediário, com comprimento mínimo maior que N_k e comprimento máximo menor que o maior comprimento N_k. Porém, essa escolha resulta em frequências mais

elevadas nunca analisadas no espectro, onde eventos transitórios podem não ser descritos (SZCZUPAK; BISCAINHO, 2009).

Para que a CQT consiga seletividade constante, é necessário que o sinal analisado permaneça estacionário ao longo de cada janela $W_{N_k}(n)$. Em sinais musicais, isso não é possível, principalmente quando a análise é realizada sobre componentes de baixa frequência. Por exemplo, uma CQT com $\beta = 1$ para análise de uma frequência de uma nota Lá com $f = 27.5Hz$ tem um intervalo de análise com duração de $612ms$. Um sinal musical estacionário por cerca de 20ms analisado pela CQT resulta em uma análise realizada sobre componentes de baixa frequência nos períodos não estacionários do sinal (SZCZUPAK; BISCAINHO, 2009).

É importante ressaltar que as transformadas DTFT, STFT e CQT, revisadas nas sub-seções anteriores, não foram explicitamente utilizadas para o desenvolvimento deste trabalho. Entretanto, foram revisadas visando formar a base para que o autor tivesse condições de desvendar uma ferramenta matemática mais ampla e que foi diretamente utilizada: a *Transformada Wavelet*.

3.6 Transformada *Wavelet*

A primeira menção sobre *wavelets* aparece na tese de doutorado de Alfred Haar em 1909, sendo a primeira literatura conhecida na área (OLIVEIRA, 2007a). Mas foi no início da década de 80 em que ela passou a ter uma identidade própria. Os franceses Jean P. Morlet e Alex Grossmann introduziram o conceito de *wavelets*, originárias de estudos de curta duração associada a pacotes de ondas acústicas sísmicas. O termo original “ondellettes”, significa algo como “pequenas ondas”, e *wavelets* correspondem a uma versão anglofônica. Alguns anos depois, as *wavelets* de Morlet atraíram a atenção do matemático Yves Meyer, que ajudou a enriquecer e amadurecer a nova teoria, com paralelos surpreendentes com diversos outros campos da matemática (GALVÃO et al., 2001). Yves Meyer construiu uma das primeiras *wavelets* não triviais, e Ingrid Daubechies construiu o mais usado conjunto de *wavelets* ortogonais. Em 1989, Stéphane Mallat, um estudante de processamento de imagens, estabeleceu a ligação da teoria *wavelet* com o processamento digital de sinais, desenvolvendo um algoritmo para calcular *wavelets* de forma computacionalmente eficiente.

A teoria *wavelet* se desenvolveu nos campos da matemática, engenharia, física quântica, e hoje tem se proliferado em uma larga gama de aplicações, incluindo: visão computacional e humana, radar, computação gráfica, hidrodinâmica, astronomia, predição de terremotos e maremotos, turbulência, descontaminação de sinais, estatística, análise de sinais médicos, processamento de voz, modelagem de sistemas lineares, modelagem geométrica, análise de transitório e falhas em linhas de potência, visualização volumétrica, sinais musicais, entre muitas outras, não sendo esta lista nem um pouco exaustiva (OLIVEIRA, 2007a).

Os problemas na resolução do tempo e da frequência, resultantes do fenômeno físico conhecido como o Princípio da Incerteza de Heisenberg (ver seção 2.6), indiferentes em relação à transformada usada, resultaram na criação de forma alternativa de análise, conhecida como Análise Multi-Resolucional (AMR). Como o próprio nome indica, ela analisa o sinal em frequências diferentes com diferentes resoluções, impondo-se uma alta resolução temporal e baixa resolução na frequência para frequências altas, e uma alta resolução frequencial e uma resolução temporal baixa para as frequências mais baixas (OLIVEIRA, 2007a).

A Transformada *Wavelet* (WT - *Wavelet Transform*) foi desenvolvida como uma alternativa para solucionar o problema da resolução. Ela surgiu nos últimos anos como uma ferramenta adequada para utilização na manipulação de sinais complexos não estacionários. Ao contrário de Fourier, em que a decomposição do sinal é feita em termos de senos e cossenos, as funções são localizadas no tempo e sem escala fixa (BALDISSERA; ORTH; STEMMER, 2001). As *wavelets* são em maioria de suporte compacto, i.e., restritas a um intervalo de tempo bem definido enquanto as bases de funções da Transformada de Fourier oscilam eternamente (SHIRADO et al., 2015).

A WT permite decompor um sinal em diferentes componentes frequenciais, nos quais se pode estudar cada componente separadamente em sua escala correspondente. O sinal pode ser analisado em diferentes escalas de forma independente, suprimindo ou reforçando alguma de suas características particulares (JR; DAMIANI, 2014). A WT é particularmente útil para a análise de transientes, aperiodicidades, e outras características de sinais não estacionários, onde mudanças na morfologia do sinal podem ser destacadas sobre as escalas de interesse.

As técnicas de *wavelet* dispõem de uma variedade de funções *wavelet* disponíveis, permitindo escolher uma função mais apropriada para um sinal sob análise, em contraste com a análise de Fourier que é restrita à morfologia de um recurso: a senóide. A WT é essencialmente dividida em duas distintas classes: a Transformada de *Wavelet* Contínua (CWT - *Continuos Wavelet Transform*) e a Transformada de *Wavelet* Discreta (DWT - *Discrete Wavelet Transform*) (ADDISON; WALKER; GUIDO, 2009).

3.6.1 Transformada de *Wavelet* Discreta

A DWT é uma das ferramentas mais poderosas para análise tempo-frequencial de sinais, com uma aplicabilidade extremamente relevante em várias áreas da ciência (GUIDO, 2015). Na sua forma mais comum, a DWT emprega uma grade diádica, onde a transformação integrante permanece contínua, mas os parâmetros de escala e translação são discretizados em uma grade local. O sinal de entrada é tratado como uma aproximação inicial para o sinal contínuo subjacente, onde a WT pode ser computada discretamente (ADDISON; WALKER; GUIDO, 2009).

Como não é viável calcular a WT para todos os possíveis valores de escala a e translação b no conjunto dos números reais, é comum fazer a seguinte restrição:

$$a = 2m,\ b = n2m \tag{3.22}$$

onde m e n são números inteiros, representando respectivamente, o nível de escala e o índice de translação (GALVÃO et al., 2001). A DWT pode ser expressa por:

$$\psi_{m,n}(t) = \frac{1}{\sqrt{a_0^m}}\psi\left(\frac{t - nb_0a_0^m}{a_0^m}\right),\ \text{com m e n inteiros} \tag{3.23}$$

onde os parâmetros m e n controlam a dilatação e a translação, respectivamente. O parâmetro a_0 é um passo de dilatação específico fixado em um valor maior que 1; e b_0 é o parâmetro de localização para valores maiores que zero. Em outras palavras, o sinal é amostrado escolhendo-se valores de escalas e translações baseados em potência de dois. Esta forma de escala logarítmica de ambos os passos de dilatação e translação é chamada de grade diádica (ADDISON; WALKER; GUIDO, 2009), guardando semelhanças com a notação musical, em que potências de dois estão relacionadas com intervalos (oitavas) e durações das notas (GALVÃO et al., 2001). A grade diádica

é uma maneira simples e eficiente de discretização e presta-se à construção de bases *wavelets* ortonormais. Uma base ortonormal *wavelet* é um conjunto de vetores que podem definir completamente um sinal $x(t)$, perpendiculares uns aos outros. A DWT usando uma grade diádica pode ser escrita como:

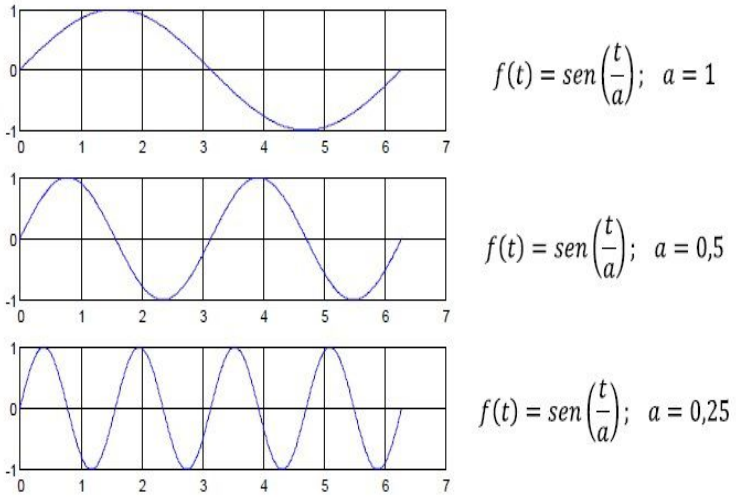
$$T_{m,n} = \int_{-\infty}^{\infty} \psi_{m,n}(t)dt, \tag{3.24}$$

sendo $T_{m,n}$ conhecido como o coeficiente *wavelet* em índice de escala e localização. A integral é contínua, porém apenas os parâmetros de escala e translação são discretizados em grade. Para a reconstrução do sinal, os coeficientes da DWT são somados ao infinito sobre m e n para obter o sinal original de volta (ADDISON; WALKER; GUIDO, 2009).

3.6.2 Função de escala e função de translação

Uma função de escala significa aplicar um alongamento ou compreensão a uma função *wavelet* (MISITI, 1997). Na Figura 19 tem-se um exemplo desse processo, em que uma senóide é modificada por um fator de escala a, progressivamente. A

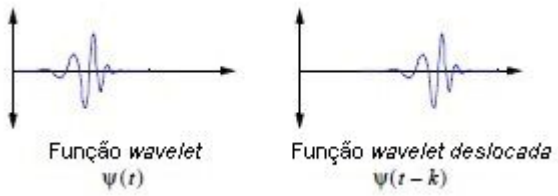
Figura 19 – Escalamento de uma senóide com fator de escala $a = 1$; $a = \frac{1}{2}$ e $a = \frac{1}{4}$.



Fonte: Misiti (1997)

função de translação significa aplicar um atraso ou adiantamento no tempo a uma função *wavelet* (MISITI, 1997). Matematicamente, atrasar uma função $f(t)$ por k é representado por $f(t-k)$. Na Figura 20 ilustra-se esse processo, em que uma senóide é modificada por um atraso no tempo.

Figura 20 – Função de translação no tempo aplicada sobre uma função *wavelet*.



Fonte: Misiti (1997)

3.6.3 Cálculo da DWT

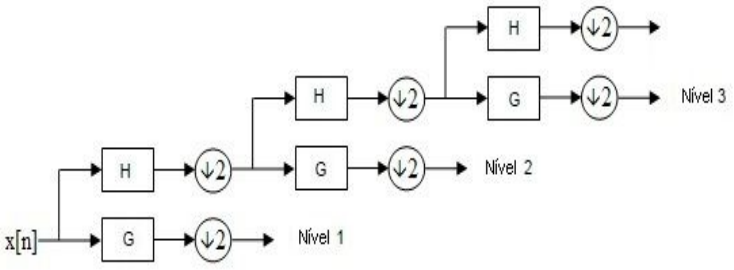
Para o cálculo da DWT de um sinal de tempo discreto, o algoritmo de Mallat é o método mais comumente utilizado. Dado um sinal no tempo discreto $f[\cdot]$ de comprimento N, o procedimento consiste na convolução de $f[\cdot]$ com duas matrizes dadas por $h[\cdot]$ e $g[\cdot]$, geralmente com um mesmo comprimento M, representando respectivamente, um filtro passa-baixa e um filtro passa-alta (GUIDO, 2015). O sinal original é decomposto em blocos baseados em potência de dois, transformando-se inicialmente em duas sub-bandas de frequência (baixa frequência e alta frequência), desmembrando o sinal em um nível de aproximação, pelo filtro passa-baixa, e um nível de detalhes, graças ao filtro passa-alta. De acordo com esse ponto de vista, a DWT de $f[\cdot]$ consiste de uma sequência de operações de filtro, à qual permite $f[\cdot]$ ser decomposta em sub-sinais com conteúdos de diferentes frequências.

O algoritmo de Mallat é implementado por meio de um banco de filtros do tipo QMF (*Quadrature Filter Mirror Pair*). Esses bancos de filtros implementam uma transformação *wavelet* ortogonal rápida que requer somente $O(N)$ operações para sinais de tamanho N (MALLAT, 1999). Os blocos resultantes da decomposição do sinal original pelos filtros passa-alta e passa-baixa representam dois sinais em que cada um contém o mesmo número de amostras do sinal original, dobrando então a frequência de amostragem (GUIDO, 2015). Um processo de sub-amostragem então é realizado, consistindo em descartar cada segunda amostra dos sinais resultantes, passando o número de amostras para cada sinal à metade, dando origem aos níveis de aproximação e detalhes. Cada nível de aproximação pode ser novamente decomposto em dois novos blocos de aproximação e detalhes, e assim sucessivamente.

Na Figura 21 é ilustrado um exemplo da decomposição de um sinal em blocos

por meio do algoritmo de Mallat (MISITI, 1997). O símbolo $(\downarrow 2)$ representa a operação de sub-amostragem, que consiste em eliminar todos os coeficientes de índice par de uma sequência, garantindo assim que o número total de pontos permaneça sempre constante.

Figura 21 – Exemplo de um sinal decomposto em níveis de aproximação e detalhes



Fonte: Suraj et al. (2014)

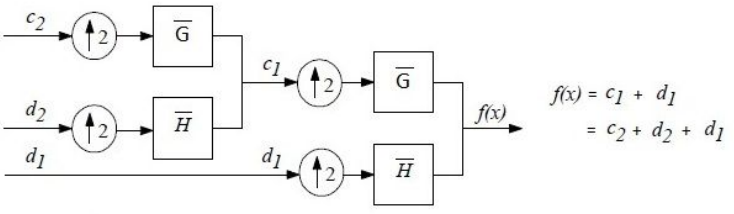
O processo iterativo do algoritmo de Mallat permite a decomposição do sinal em componentes de menor resolução com seus respectivos detalhes. A função de escala, conhecida como *wavelet* pai, e a função *wavelet*, conhecida como *wavelet* mãe, são respectivamente definidas de modo recursivo como:

$$\sum_k h_k \phi(2x - k) \tag{3.25}$$

$$\sum_k g_k \phi(2x - k). \tag{3.26}$$

O processo de reconstrução do sinal é possível pela reconstrução de seus componentes, como ilustrado na Figura 22. O símbolo $(\uparrow 2)$ representa a operação de inserção de zeros entre os pontos de uma sequência (GALVÃO et al., 2001).

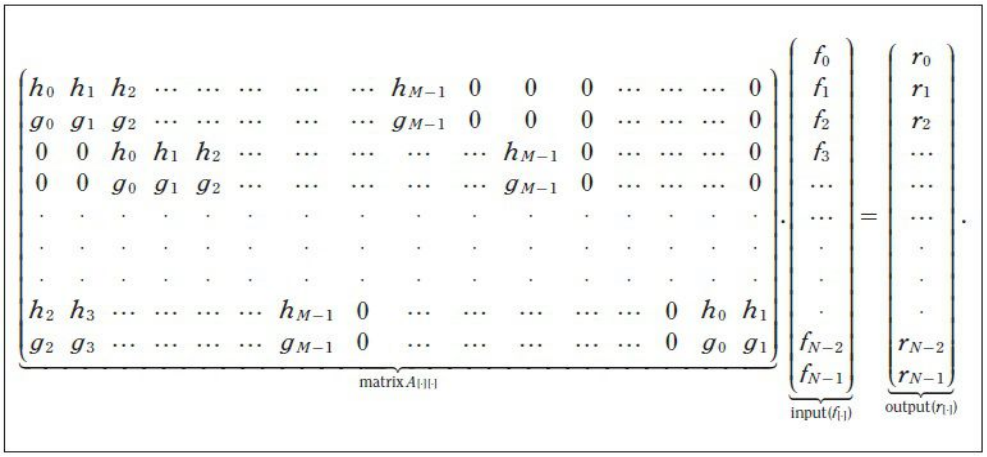
Figura 22 – Exemplo de reconstrução do sinal decomposto em dois níveis.



Fonte: Galvão et al. (2001)

O procedimento para o cálculo da DWT consiste numa multiplicação simples de $f[\cdot]$ por $h[\cdot]$ e $g[\cdot]$. Uma matriz $A[\cdot][\cdot]$, formada pelos coeficientes dos filtros avança desde o primeiro par de linhas até o último, onde uma mudança se torna necessária assim que $h[\cdot]$ e $g[\cdot]$ começam a ser escritas duas posições à frente em cada par subsequente. No caso dos coeficientes ultrapassarem o comprimento da linha, eles são empurrados de volta ao início na mesma linha. As posições restantes da matriz são prenchidas com zeros (GUIDO, 2015). A Figura 23 contém um exemplo.

Figura 23 – Matriz $A[\cdot][\cdot]$ multiplicada por um entrada de sinal discreto $f[\cdot]$.



Fonte: Guido (2015)

Calculado o resultado da convolução, a DWT de $f[\cdot]$ corresponde à concatenação do sub-sinal de aproximação de tamanho $\frac{N}{2}$ com o sub-sinal de detalhamento de tamanho $\frac{N}{2}$, resultando em uma saída $r[\cdot]$ de mesmo tamanho N do sinal de entrada. Este resultado é conhecido como primeiro nível. Se o sinal de detalhe é mantido intacto e o sinal de aproximação é servido como uma entrada nova para o algoritmo, dois sub-sinais de tamanho $\frac{N}{4}$ são obtidos. O processo pode ser repetido até que o comprimento de aproximação se torne igual a 1.

Com os avanços da decomposição do sinal, cada aproximação é reduzida à metade. Se a dimensão da entrada torna-se menor do que os filtros, as matrizes podem não ser compatíveis para realizar a multiplicação, sendo necessário repetir a entrada quantas vezes forem necessárias para tornar possível efetuar as multiplicações (GUIDO, 2015). Na equação 3.27 tem-se um exemplo de uma matriz de filtros $A[\cdot][\cdot]$ com duas linhas e quatro colunas, e um sub-sinal como entrada. A repetição das

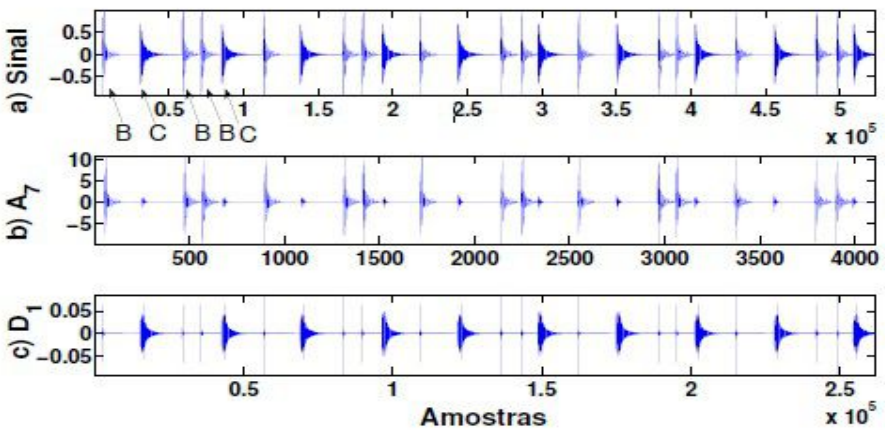
duas primeiras linhas é necessária para efetuar a multiplicação.

$$\underbrace{\begin{bmatrix} \frac{1+\sqrt{3}}{4\sqrt{2}} & \frac{3+\sqrt{3}}{4\sqrt{2}} & \frac{3-\sqrt{3}}{4\sqrt{2}} & \frac{1-\sqrt{3}}{4\sqrt{2}} \\ \frac{1-\sqrt{3}}{4\sqrt{2}} & \frac{-3+\sqrt{3}}{4\sqrt{2}} & \frac{3+\sqrt{3}}{4\sqrt{2}} & \frac{-1-\sqrt{3}}{4\sqrt{2}} \end{bmatrix}}_{matrizA[\cdot][\cdot]} \begin{bmatrix} \frac{156+28\sqrt{3}}{32} \\ \frac{164+28\sqrt{3}}{32} \\ \frac{156+28\sqrt{3}}{32} \\ \frac{164+28\sqrt{3}}{32} \end{bmatrix} \begin{matrix} \} \text{ entrada} \\ \\ \} \text{repetição} \end{matrix} \tag{3.27}$$

A maior parte das aplicações baseadas na DWT também exigem uma transformada inversa, ou seja, reverter o processo para a reconstrução do sinal. Aplicações envolvendo análises de imagens e compreensão de áudio são exemplos. O processo de inversão de uma DWT é conhecido como Transformada de *Wavelet* Discreta Inversa (IDWT - *Inverse Discrete Transform Wavelet*) (GUIDO, 2011). O sinal reconstruído assume uma forma que corresponde a uma combinação das formas das funções de escala e *wavelet* associadas com os filtros.

A DWT permite uma análise de sinais musicais com grande precisão, devido a ser localizada nas dimensões de tempo e frequência. Ela pode fornecer resultados interessantes para detecção de periodicidades, detecção de frequência fundamental, remoção de ruídos entre outras (JR; DAMIANI, 2014). Na Figura 24 tem-se um exemplo de DWT aplicada sobre um sinal de bateria formado por um bumbo(B) e uma caixa(C), decompostas em 7 níveis usando a base Coiflet. Em 24(a) encontra-se o sinal original seguido por 24(b) e 24(c), representando o nível de aproximação A_7, e o nível de detalhe D_1.

Figura 24 – Decomposição *wavelet* em 7 níveis coiflet. (a) sinal original; (b) nível A7 de aproximação; (c) nível D1 de detalhamento.



Fonte: Jr e Damiani (2014)

O compasso repete um padrão (B, C, B, B, C) por quatro vezes seguidas, e observando o nível de aproximação A_7, o padrão de frequência predominante revelado é do bumbo. O nível de detalhamento D_1 contém uma presença mais marcante da caixa. Pode-se concluir que a resolução promovida pela DWT separou o sinal em diferentes escalas.

3.6.4 Famílias da Transformada *Wavelet*

A DWT possui uma ampla variedade de funções *wavelets*, em que cada função apresenta características próprias e diferentes suportes para cada filtro (SHIRADO et al., 2015). Na Tabela 1 resume-se de maneira geral as particularidades de cada uma das principais famílias de filtros *wavelets*. Na Figura 25, ilustra-se um exemplo

Tabela 1 – Famílias *wavelets*

Família	Suporte	Fase	Observação
Haar	2	Linear	é a mais simples das *wavelets* criada por Alfred Haar
Daubechies	par, maior que 4	não linear	resposta ao impulso *maximally flat*, criada por Ingrid Daubechies
Symmlets	par, múltiplo de 8	não linear	resposta ao impulso mais simétrica
Coiflets	Par múltiplo de 6	quase linear	resposta ao impulso quase simétrica, criada por Ronald Coifman
Beylkin	18	não linear	otimizada para áudio em geral

Fonte: Shirado et al. (2015)

da *wavelet* de Daubechies de ordem 4 (4 coeficientes), com sua função de escala e função *wavelet* correspondente. Seus coeficientes podem ser definidos por:

$$(h_0, h_1, h_2, h_3) = \Big(\frac{1+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{3-\sqrt{3}}{4\sqrt{2}}, \frac{1-\sqrt{3}}{4\sqrt{2}}\Big). \tag{3.28}$$

A partir dos coeficientes pode-se construir a função escala

$$\phi t = \sqrt{2} \sum_{k=0}^{2N-1} h_k \phi(2t-k) \tag{3.29}$$

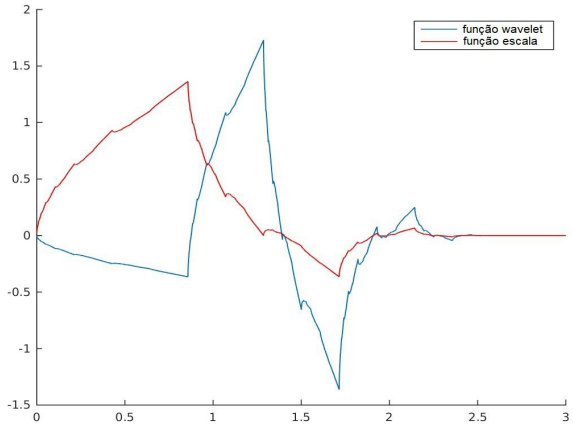
e calcular os coeficientes g_n

$$(g_0, g_1, g_2, g_3) = \Big(\frac{1-\sqrt{3}}{4\sqrt{2}}, \frac{-3+\sqrt{3}}{4\sqrt{2}}, \frac{3+\sqrt{3}}{4\sqrt{2}}, \frac{-1-\sqrt{3}}{4\sqrt{2}}\Big) \tag{3.30}$$

resultando na função *wavelet* dada por

$$\psi t = \sqrt{2} \sum_{k=0}^{2N-1} g_k \phi(2t - k). \tag{3.31}$$

Figura 25 – Função escala e função *wavelet* de Daubechies de ordem 4.



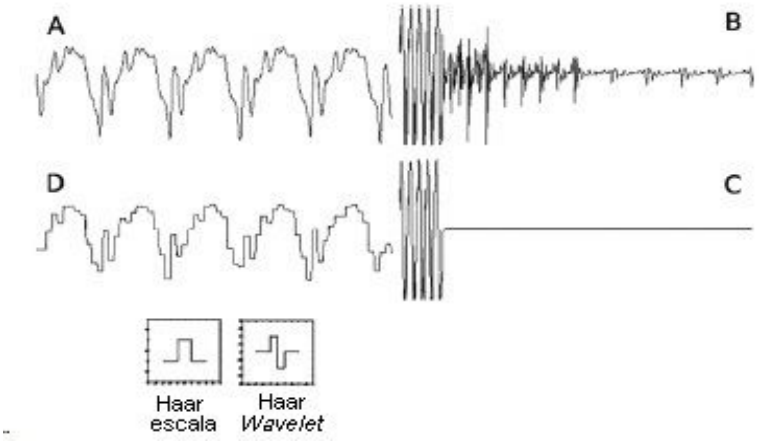
Fonte: Daubechies e Paul (1987)

As funções de escala e *wavelets* são importantes quanto à sua escolha para a análise do sinal. Se a aplicação requer a reconstrução do sinal, as formas associadas aos filtros podem ser consultadas. Isto se deve à modificação do sinal transformado, onde quanto mais o sinal é modificado antes da reconstrução, mais o sinal reconstruído assume uma forma correspondente as funções de escalonamento e *wavelets* (GUIDO, 2011). Na Figura 26 tem-se um sinal de amostra, a sua DWT realizada com um filtro de Haar, a modificação introduzida no sinal transformado e, por fim, a sua IDWT. Pode-se notar que as formas dos filtros de Haar são claramente visíveis no sinal reconstruído.

A escolha de uma função *wavelet* é um dos primeiros problemas práticos que devem ser abordados para encontrar uma boa análise *wavelet* de uma série temporal. Uma escolha razoável depende muito da aplicação em questão, uma vez que existe uma interação apreciável entre o valor desejado e as propriedades que precisamos em uma função *wavelet* para atingir esse objetivo (GANCHEV et al., 2014). Propriedades como resposta em frequência, regularidade ou suavidade podem ser determinantes na escolha da função *wavelet* a ser usada.

Frequentemente as *wavelets* são classificadas em famílias de acordo com o número de momentos nulos (*vanishing moments*). Um sinal pode ser aproximadamente

Figura 26 – Utilizando filtros de Haar para análise e síntese de um sinal. (a) sinal original; (b) terceiro nível da DWT; (c) modificação do terceiro nível; (d) sinal reconstruído através da combinação de dilatações e translações. Abaixo as funções de escala e *wavelet*.



Fonte: Guido (2011)

descrito por um polinômio de grau menor que M e por uma *wavelet* que possui M momentos nulos se seus coeficientes de detalhamento são aproximadamente zero. O M-ésimo momento pode ser calculado por

$$\sum_{k=0}^{p-1} {t_k}^m \psi_(t_k) \tag{3.32}$$

sendo p a quantidade de coeficientes da função *wavelet* ψ, m o momento desejado e t cada coeficiente onde a função pode assumir valores diferentes de zero (JENSEN; COUR-HARBO, 2001).

A *wavelet* é dita ser de suporte compacto se a maioria da energia desta *wavelet* está restrita a um intervalo finito, ou seja, se a função é exatamente zero quando o tempo ou a frequência vão a infinito (GANCHEV et al., 2014). Portanto, os comprimentos tanto da *wavelet* como dos filtros de escala são responsáveis pela seletividade de frequência e pela resolução temporal. Um elevado número de momentos nulos permite melhor compressão das partes regulares do sinal aumentando também o tamanho do suporte das *wavelets*. Isso pode resultar em equações mais complexas e computação mais lenta. Funções com suporte maior aumentam a resolução em frequência, mas diminuem a resolução temporal do sinal transformado. A escolha da *wavelet* deve levar em consideração um suporte que não comprometa nenhuma das resoluções (SOUZA et al., 2007).

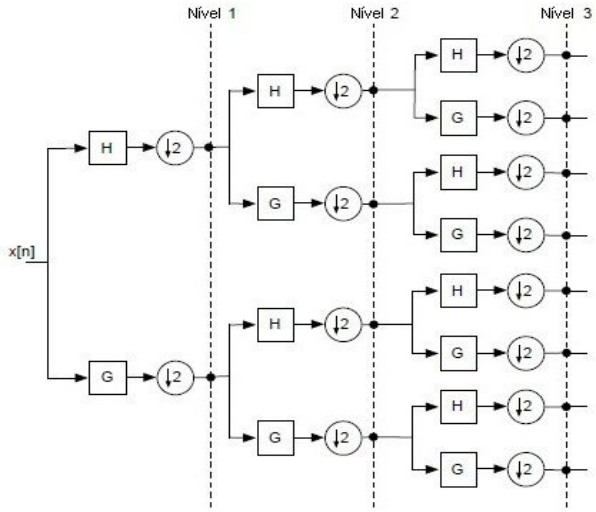
A regularidade ou suavidade de uma *wavelet* são determinadas pelo seu número de momentos nulos. Quanto maior o número de momentos, maior será o número de coeficientes e mais suave será a *wavelet*, com maior probabilidade de reconstrução perfeita do sinal decomposto pela transformada. Para aplicações que não necessitam da IDWT, a decisão pode ser baseada na resposta de frequência e fase de $h[\cdot]$ e $g[\cdot]$. Diferentes *wavelets* resultarão em diferentes filtros com diferentes respostas em frequências (GUIDO, 2011). Uma *wavelet* de ordem maior oferece uma melhor resolução frequencial, exibindo respostas de fase linear e maximamente planas em suas bandas de passagem e corte.

3.7 Transformada *Wavelet-Packet*

Uma generalização da DWT foi proposta por Coifman e Wickerhauser no ano de 1992, conhecida como Transformada *Wavelet-Packet* (WPT - *Wavelet Packet Transform*). Apresentanto uma complexidade de tempo igual a $O(nlogn)$ (MALLAT, 1999), a WPT busca refinar a decomposição do sinal para todas as faixas de frequência, onde tanto os coeficientes de aproximação quanto os coeficientes de detalhes são subdivididos. Ao contrário da DWT que recursivamente decompõe apenas a sub-banda passa-baixa, a WPT decompõe as duas sub-bandas em cada nível sucessivamente, fornecendo um melhor controle de resolução frequencial para a decomposição do sinal (GRIMALDI; CUNNINGHAM; KOKARAM, 2003). O resultado (Figura 27) é uma árvore binária onde cada nó representa um subespaço, cujas funções são divididas por um banco de filtros de dois canais, contendo o sinal aproximado em diferentes resoluções (OLIVEIRA; FALK; TÁVORA, 2017).

A WPT permite alicerçar o espaço de frequência em um número discreto de intervalos. Para a análise musical, essa possibilidade permite definir as sub-bandas correspondentes a oitavas musicais e notas musicais. Considerando apenas as frequências correspondentes às notas musicais, a caracterização do espectro se torna uma tarefa relativamente fácil (GRIMALDI; CUNNINGHAM; KOKARAM, 2003). Uma desvantagem na utilização da WPT é que se torna impossível definir um nível de decomposição único adequado tanto para a extração de tempo quanto de frequência.

Figura 27 – Exemplo de uma WPT implementada por banco de filtros.



Fonte: Marchi et al. (2014)

3.8 Redes Neurais Artificiais

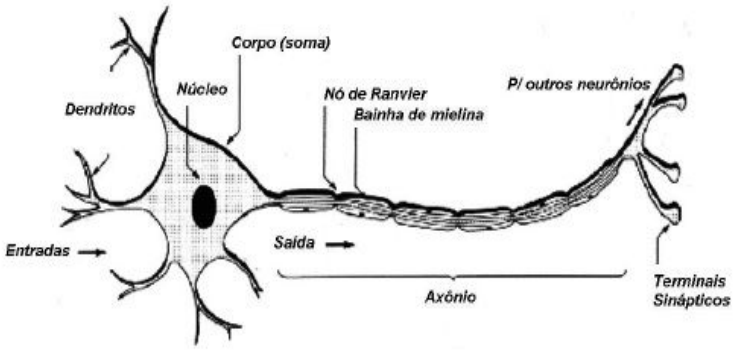
A utilização de sistemas baseados em inteligência artificial para a análise de sinais de áudio é proposta por diversos pesquisadores, como forma de permitir maior flexibilidade na exploração de suas características. As Redes Neurais Artificiais (RNAs) vêm sendo usadas com sucesso no reconhecimento de padrões há muitos anos, pois possuem a capacidade de aprendizagem a partir de um conjunto de treinamento adequado, ou seja, conseguem produzir saídas adequadas para entradas que não pertencem ao conjunto de treinamento.

A neurociência é o estudo do sistema nervoso e suas funcionalidades, além de estruturas e processos de desenvolvimento. A medição da atividade do cérebro teve início em 1929, por Hans Berger com a invenção do eletroencefalógrafo (EEG), e mais recentemente pelo processamento de imagens por ressonância magnética funcional, dando aos neurocientistas imagens sem precedentes de detalhes da atividade cerebral (RUSSELL; NORVIG, 2004). As medições são ampliadas por avanços na gravação da atividade dos neurônios em uma única célula, podendo estes serem estimulados eletricamente, quimicamente, ou opticamente, permitindo que seus relacionamentos neuronais de entrada e saída sejam mapeados. Basicamente, o processamento de informações é distribuído por meio de camadas de neurônios, sendo que todos os neurônios dentro dessas camadas processam as suas entradas simultaneamente e

independentemente (FARIAS et al., 2009).

Um neurônio biológico é uma unidade de processamento de informação fundamental para a operação de uma rede neural (HAYKIN, 2001). Ele possui basicamente várias ramificações chamadas de dendritos, responsáveis por receberem sinais de outros neurônios interligados, e um ramo principal chamado axônio que, na sua outra extremidade, forma sinapses, ou ligações, com os dendritos dos outros neurônios. Quando impulsos combinados excedem um determinado limiar, o neurônio dispara um impulso ao longo do axônio (FARIAS et al., 2009). Essas características capturam o essencial para os modelos neurais de computação. Na Figura 28, tem-se o esquema de um neurônio biológico.

Figura 28 – Neurônio biológico.

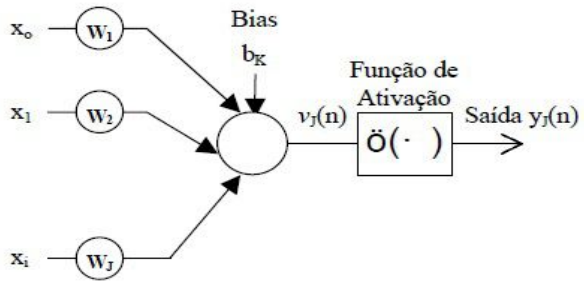


Fonte: Farias et al. (2009)

Desde 1943, têm sido desenvolvidos modelos muito mais detalhados e realistas para os neurônios, levando ao campo moderno da neurociência computacional (RUSSELL; NORVIG, 2004). O momento fundador é atribuído ao cientista neurofisiólogo Warren Sturgis McCulloch (1899-1969) e ao lógico Walter Pitts (1923-1969) ao publicar em 1943 o artigo clássico *"A logical calculus of the ideas immanent in nervous activity"* onde propunham um modelo simples de rede do tipo neuronal capaz de realizar operações lógicas (QUINTAIS, 2009). O modelo consiste numa rede de unidades de processamento que podem ser ativadas ou inibidas no seu funcionamento quando a soma ponderada de suas entradas atinge um certo limiar. Uma analogia do funcionamento do neurônio biológico é mostrado na Figura 29.

Os sinais de entrada $x_o, x_1, ..., x_i$ são também denominados estímulos, ou seja,

Figura 29 – Neurônio artificial.



Fonte: Moutinho e Neto (2002)

equivalentes matemáticos aos mais variados sinais que são recebidos pelo cérebro humano, como audição, tato, olfato e outros. Cada neurônio recebe todo o conjunto de entradas x ponderadas por um conjunto de pesos W_J. Todas as entradas são multiplicadas por seus devidos pesos e são somadas a um valor fixo externo b_k chamado bias. A expressão final V_J é então aplicada à função de ativação, resultando na saída Y_J, eventualmente podendo ser aplicada à entrada de um ou mais neurônios (MOUTINHO; NETO, 2002). Em termos matemáticos, um neurônio k pode ser descrito pelo seguinte par de equações:

$$u_k = \sum_{j=1}^{j} w_{kj} x_j \tag{3.33}$$

$$v_k = (u_k + b_k). \tag{3.34}$$

$$y_k = \varphi(v_k). \tag{3.35}$$

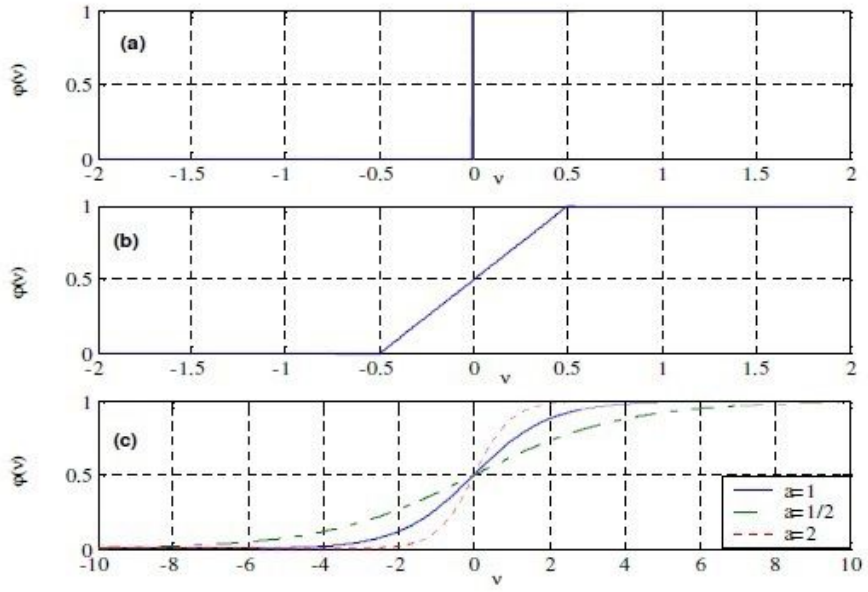
Cada unidade k primeiro calcula uma soma ponderada de suas entradas, onde $x_1, x_2, ..., x_i$ são sinais de entrada; $w_{k1}, w_{k2}, ..., w_{kn}$ são os pesos sinápticos e u_k é a saída do combinador linear (HAYKIN, 2001). Em seguida, aplica-se uma função de ativação g a essa soma para obter a saída, onde: b_k é o bias, φ é a função de ativação e y_k é o sinal de saída do neurônio.

As vantagens das redes neurais se evidenciam na sua generalização, ou seja, a capacidade de apresentar saídas coerentes para entradas que não estavam presentes durante o processo de aprendizagem. As RNAs também possuem outras potencialidades, como a capacidade de trabalhar com problemas não-lineares, sua adaptabilidade e tolerância a falhas (HAYKIN, 2001).

3.8.1 Funções de ativação

A função de ativação é responsável pelo valor da saída em função dos valores locais. É ela que introduz o comportamento não-linear do neurônio e produz o estado ligado/desligado semelhante a um neurônio natural (FARIAS et al., 2009). Existem três tipos básicos de função de ativação (HAYKIN, 2001), conforme podem ser vistas na Figura 30 e descritas a seguir.

Figura 30 – Funções: (a) de limiar; (b) linear por partes e (c) sigmóide.



Fonte: Haykin (2001)

Função de limiar: utilizada no neurônio de McCulloch-Pitts, definida por:

$$\varphi(v_k) = \begin{cases} 1, \text{se } v_k \geq 0, \\ 0, \text{se } v_k < 0, \end{cases} \tag{3.36}$$

onde v_k é o campo induzido do neurônio 3.34. A saída de um neurônio assume o valor 1 se o campo induzido é não-negativo, e 0 caso contrário;

Função Linear por Partes: é definida por:

$$\varphi(v) = \begin{cases} 1, \text{se } v \geq \frac{1}{2}, \\ v, \text{se } 1/2 > v > -\frac{1}{2}, \\ 0, \text{se } v \leq -\frac{1}{2}, \end{cases} \tag{3.37}$$

onde se assume que o fator de amplificação dentro da região linear é a unidade;

Função Sigmóide: é a função mais utilizada, definida por:

$$\varphi(v) = \frac{1}{1 + exp(-av)} \tag{3.38}$$

onde a é o parâmetro de inclinação da função. Variando-se a, obtemos diferentes inclinações como ilustrado na Figura 30(c).

As funções de ativação descritas se estendem em um intervalo de 0 a +1. Para casos em que é necessário que o resultado fique entre -1 e +1, a função de ativação utilizada deve ser ímpar (HAYKIN, 2001). Especificamente, para a função sigmóide, sua correspondente ímpar é a *função tangente hiperbólica*, definida na equação 3.39.

$$\varphi(v) = tanh(v) \tag{3.39}$$

3.8.2 Processo de aprendizagem

No processo de aprendizagem, os pesos sinápticos e níveis de *bias* da rede são modificados de forma a alcançar um objetivo predefinido (FARIAS et al., 2009). Uma base de exemplos é necessária para possibilitar um processo iterativo no qual a rede ajusta os pesos sinápticos, adaptando-se aos casos de entrada. O “aprendizado” surge da adaptação de pesos da rede, que deverá aprender a responder aos estímulos de entrada de acordo com os exemplos que foram apresentados.

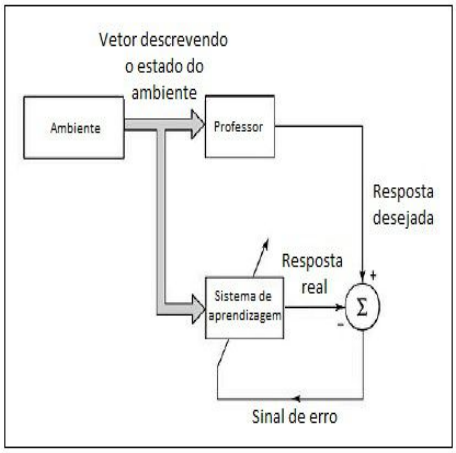
A mudança nos parâmetros da rede resulta na mudança do seu comportamento geral. Ao invés de especificar todos os detalhes de uma computação para cada caso, tem-se a possibilidade de treinar uma rede em âmbito mais genérico para fazer esta computação (HAYKIN, 2001). Não há um algoritmo de aprenzagem único para o projeto de RNAs. Eles diferem pela forma como é formulado o ajuste dos peso sinápticos. Os procedimentos que levam uma RNA a aprender determinadas tarefas podem ser de dois tipos: aprendizado supervisionado e aprendizado não-supervisionado.

3.8.3 Aprendizado supervisionado

Na Figura 31, o treinamento supervisionado está representado. O conhecimento é representado por um conjunto de exemplos de entrada e saída, chamados de

pares de treinamento (HAYKIN, 2001). O ambiente é desconhecido pela RNA, onde consideramos o “professor” como tendo conhecimento sobre o ambiente. Em virtude desse conhecimento prévio, o professor é capaz de fornecer à rede uma resposta desejada para uma determinada entrada de treinamento.

Figura 31 – Diagrama em blocos da aprendizagem com um professor.



Fonte: Haykin (2001)

A saída da rede é calculada e comparada com o correspondente vetor objetivo, fornecido pelo professor. Um sinal de erro é definido como a diferença entre a resposta desejada e a resposta real da rede. Esse sinal é então realimentado e os pesos são atualizados de acordo com um algoritmo de aprendizagem determinado, a fim de minimizar o erro. O ajuste é realizado iterativamente até que obtenham níveis de erros mais baixos.

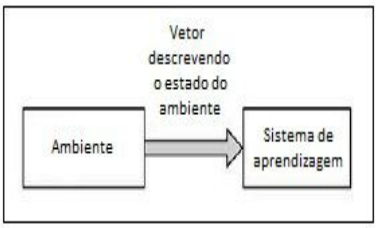
O objetivo da RNA é emular o professor, transferindo o seu conhecimento da forma mais completa possível para a entidade neural, permitindo à rede lidar com o ambiente por si mesma, sem ajuda do professor (HAYKIN, 2001). Um sistema com um conjunto adequado de exemplos de entrada-saída e tempo suficiente para realizar o treinamento é normalmente capaz de realizar classificações de padrões e aproximação de funções.

3.8.4 Aprendizado não-supervisionado

Ao contrário do supervisionado, no aprendizado não-supervisionado não há um professor para supervisionar o processo, como ilustrado na Figura 32. Os pesos

da rede são modificados pelo conjunto de treinamento, de forma a produzir saídas consistentes. Uma vez que a rede tenha se ajustado às regularidades estatísticas da entrada, ela desenvolve a habilidade de criar novas classes formando representações internas para codificar as características da entrada (HAYKIN, 2001).

Figura 32 – Diagrama em blocos da aprendizagem não-supervisionada.



Fonte: Haykin (2001)

3.8.5 Arquitetura das redes neurais

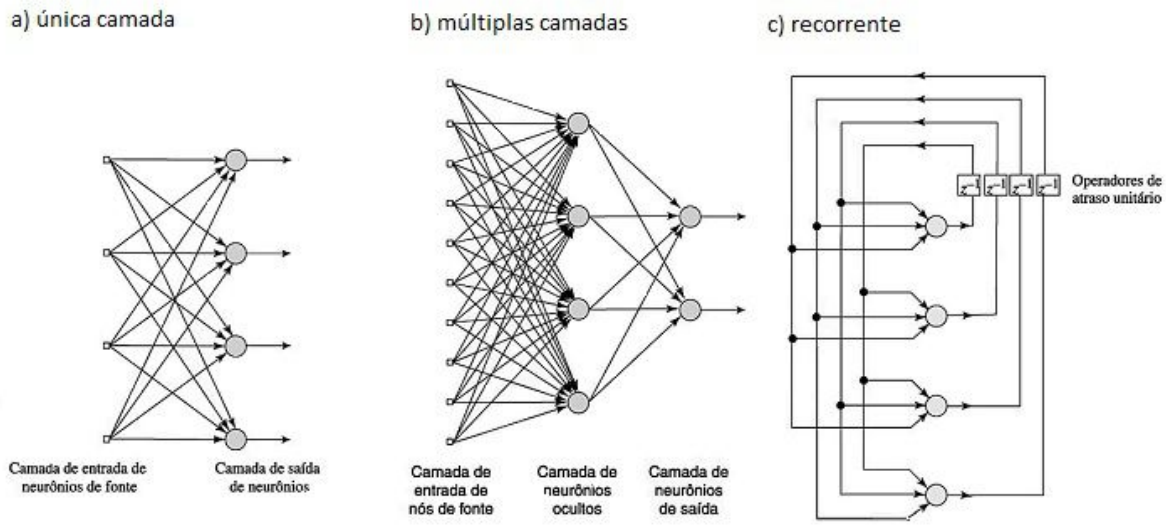
A maneira pela qual os neurônios de uma RNA estão estruturados está diretamente ligada com o algoritmo de aprendizagem usado no treinamento da rede (HAYKIN, 2001). Em geral, fundamentalmente existem três classes de arquiteturas de redes diferentes: as redes de camada única, as de múltiplas camadas, e as redes recorrentes.

As redes de camada única apresentam uma camada de entrada de nós que se projeta sobre uma camada de nós de saída, ou seja, todas as entradas estão conectadas diretamente com as saídas. A designação de camada única refere-se à camada de saída dos neurônios (HAYKIN, 2001). Sua vantagem é a capacidade de representar uma função boleana bastante complexa de forma mais compacta. Sua desvantagem está na incapacidade dos neurônios aprenderem funções simples como uma operação XOR, em razão de uma única unidade de limiar (RUSSELL; NORVIG, 2004).

As redes de múltiplas camadas se distinguem pela presença de uma ou mais camadas ocultas, responsáveis por intervir entre a camada de entrada e a camada de saída da rede. Devido ao conjunto extra de conexões da camada oculta, a rede adquire uma perspectiva global capaz de extrair estatísticas de ordem mais elevada, ideal para problemas mais complexos. As redes recorrentes se distinguem por ter laços de realimentação, onde cada neurônio pode, por exemplo, alimentar seu sinal de saída

de volta para as entradas de todos os outros neurônios. Na Figura 33, ilustram-se as redes de camada única, as de múltiplas camadas e as redes recorrentes (HAYKIN, 2001).

Figura 33 – Classes de arquiteturas de redes neurais.



Fonte: Haykin (2001)

As redes neurais de uma só camada são capazes de resolver apenas problemas linearmente separáveis, ou seja, que podem ser satisfeitos por uma reta ou hiperplano como fronteira de decisão. Seus algoritmos de treinamento utilizados ajustam os pesos de somente uma camada. Os problemas de classificação conhecidos como não-lineares exigem a utilização de algoritmos que ajustem os pesos de múltiplas camadas (BISHOP, 1995).

3.8.6 Redes *Perceptron*

O *perceptron* é a forma mais simples de uma rede neural usada para classificação de padrões linearmente separáveis, ou seja, padrões que se encontram em lados opostos do hiperplano (HAYKIN, 2001). Consiste em um único neurônio construído em torno do modelo de McCulloch-Pitts, com pesos sinápticos ajustáveis e *bias*. As redes *perceptrons* construídas com camada única limitam-se a realizar classificações com apenas duas classes de padrões linearmente separáveis.

Uma rede constituída de uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída é normalmente chamada de redes *Multi-Layer Percetron* (MLP) (HAYKIN, 2001). O sinal de entrada se propaga por meio da rede, camada por camada, produzindo um conjunto de saídas. Durante a propagação, os pesos sinápticos são todos fixos, sendo ajustados apenas na chamada retropropagação, de acordo com uma regra de correção de erro. O processo é repetido iterativamente até que o erro esteja abaixo de um limiar. Este algoritmo é popularmente conhecido como algoritmo de retropropagação de erro (*error backpropagation*). A resposta real da rede é subtraída de uma resposta desejada para produzir um sinal de erro, propagado para trás contra a direção das conexões sinápticas. A complexidade computacional do algoritmo de retropropagação é linear em relação ao seu total W de pesos sinápticos, isto é, $O(W)$.

As redes MLP são os modelos de RNAs mais utilizados (HAYKIN, 2001). Um dos problemas que ocorre durante seu treinamento é o chamado sobre-ajuste (*overfitting*). A rede memoriza os exemplos do treinamento, mas não aprende a generalizar situações novas. Existem vários métodos que podem amenizar esse tipo de problema.

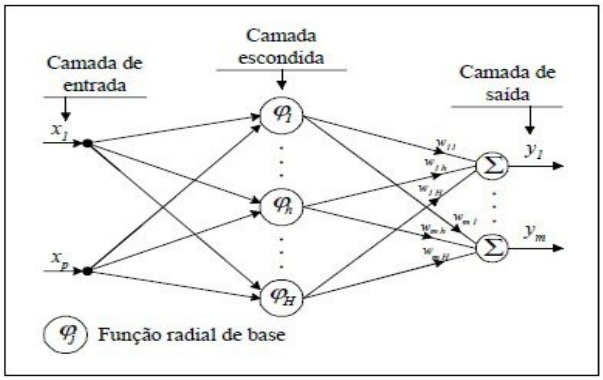
Assim como no caso das transformadas DTFT, STFT e CQT, as quais foram revisadas para consubstanciar o entendimento da DWT, as redes MLP não foram diretamente utilizadas neste trabalho. Entretanto, a breve revisão realizada sobre elas visou fornecer base para a estrutura neural usada: as redes com função de base radial.

3.8.7 Redes de Funções Radiais de Base

As Redes de Funções Radiais de Base (RBF - *Radial-Basis Function*) têm sido, recentemente, relevantes dentro do domínio das RNAs. A simplicidade do processo de treinamento e sua eficiência computacional têm sido um atrativo para diversos pesquisadores (FERNANDES; NETO; BEZERRA, 1999). São redes de alimentação para diante (*feed-forward*), inerentemente acíclicas, ou seja, o sinal é propagado somente da entrada para a saída da rede. Em sua forma mais básica, uma rede RBF envolve três camadas, conforme a Figura 34: uma camada de entrada constituída por nós que se conectam ao ambiente, uma camada oculta que aplica uma transformação não-linear do espaço de entrada para o espaço oculto e uma camada de saída linear

que fornece uma reposta ao padrão de ativação aplicado à camada de entrada. Os neurônios da camada escondida são funções radiais de base (HAYKIN, 2001).

Figura 34 – Rede neural do tipo RBF.



Fonte: Fernandes, Neto e Bezerra (1999)

As unidades da camada escondida recebem um vetor x de entrada, n-dimensional e real, e processam-no de acordo com uma função de base radial f_i. Uma rede RBF com m entradas e n saídas implementa o mapeamento $f_i : \mathbb{R}^m \rightarrow \mathbb{R}^n$, que pode ser definido por:

$$f_i(x) = f_i\left(\frac{||x - c_i||}{d_i}\right) \tag{3.40}$$

onde $c_i \in \mathbb{R}^n$ é o centro da RBF, $f_i d_i \in \mathbb{R}$ é o fator de escala para o raio $||x - c_i||$, e $||\ ||$ é tipicamente a norma euclidiana sobre $c_i \in \mathbb{R}^n$ (HASSOUN, 1995). O argumento da função de ativação de cada unidade da camada oculta calcula a distância euclidiana entre o vetor de entrada e o centro daquela unidade, ao contrário da MLP na qual se calcula o produto interno do vetor de entrada e do vetor de pesos sinápticos da unidade (HAYKIN, 2001).

As funções de base radial são classes especiais de funções cujo valor de saída aumenta ou diminui em relação à distância de um ponto central. Elas produzem um resposta significativa, diferente de zero, somente quando o padrão de entrada está dentro de uma região pequena localizada no espaço de entrada (FERNANDES; NETO; BEZERRA, 1999). Existem diferentes funções para a construção de redes RBF, porém as mais comuns são a função gaussiana, a função multiquadrática e a função *thin-plate-spline*, definidas respectivamente pelas equações 3.41, 3.42 e 3.43, onde $v = ||x - \mu||$ é a distância euclidiana, x é o registro de entrada, e μ e σ representam o centro e a

largura da função radial (BRAGA; CARVALHO; LUDERMIR, 2000).

$$\varphi(u) = e^{\left(\frac{-v^2}{2\sigma^2}\right)} \tag{3.41}$$

$$\varphi(u) = \sqrt{(v^2 + \sigma^2)} \tag{3.42}$$

$$\varphi(u) = v^2 \ln(v^2) \tag{3.43}$$

As redes RBF são apropriadas para realizar aproximações que incluem os problemas de classificação de padrões. Quatro parâmetros podem controlar o grau de precisão das aproximações: o tipo de função de base radial, o número de neurônios da camada escondida, a localização dos centros e dos raios e o ajuste dos pesos das conexões da camada de saída (HASSOUN, 1995). A aprendizagem pode ser supervisionada, não-supervisionada ou híbrida. O treinamento híbrido ocorre em dois estágios de ajuste: primeiro, são determinados os parâmetros das funções de base, ou seja, não-supervisionado, e em seguida são determinados os pesos da camada de saída de modo supervisionado. A forma das funções de base é escolhida a priori, de modo que ela tenha um comportamento adequado ao problema de regressão: a sua resposta deve decrescer monotonicamente com a distância em relação a um ponto central.

Existem duas razões principais que justificam o uso de uma rede RBF. A primeira é que uma rede RBF pode realizar classificações do tipo não-lineares, superando o problema do *perceptron* de uma camada. Uma segunda vantagem é que seu processo de treinamento é bem mas rápido que da MLP, em virtude da necessidade da MLP aplicar uma retropropagação de erro (FERNANDES; NETO; BEZERRA, 1999).

4 A ABORDAGEM PROPOSTA

Este trabalho objetivou desenvolver um sistema capaz de processar um sinal musical e identificar padrões de notas acústicas, com possíveis aplicações que visam auxiliar estudantes no aprendizado musical. Para realização do processo, foi utilizada a Transformada Wavelet-Packet (WPT) em conjunto com a técnica de autocorrelação, seguida pela classificação dos eventos por meio de uma rede neural do tipo RBF. Na Figura 35 consta a sequência de etapas seguidas: a aquisição do sinal de áudio, aplicação da WPT, aplicação da autocorrelação e, por fim, classificação dos eventos pela rede neural, identificando-se as notas musicais de acordo com a saída do algoritmo inteligente.

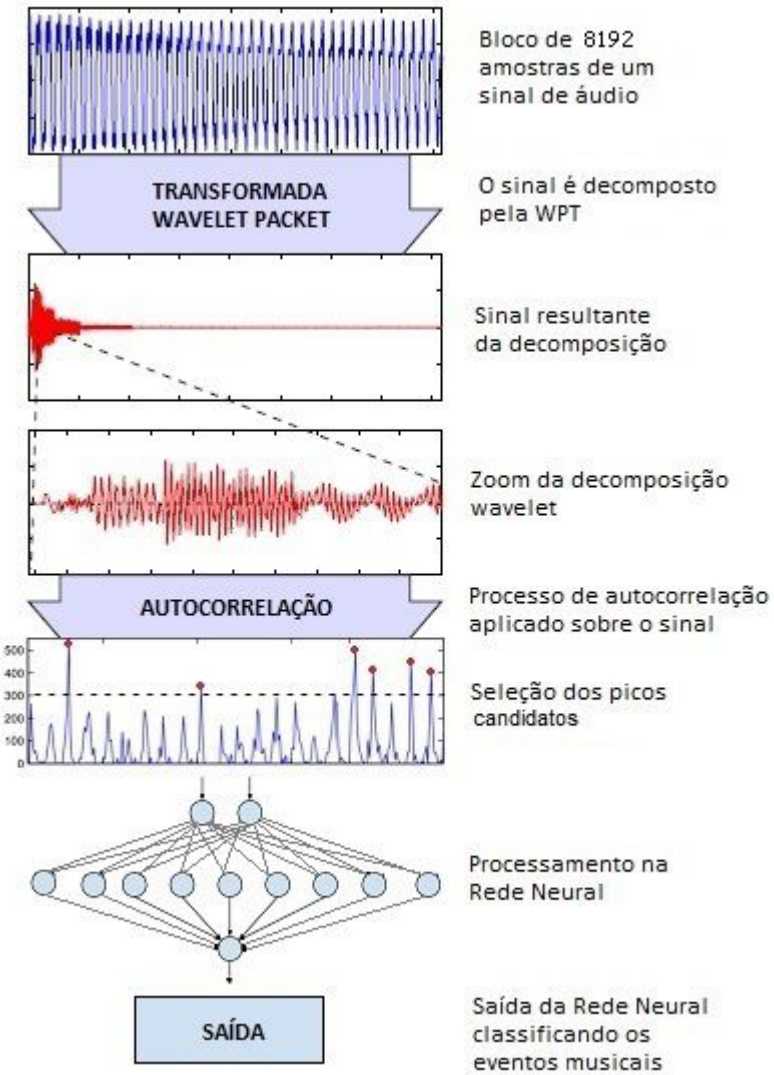
4.1 Metodologia e implementação

Este trabalho fundamentou-se na aplicação da análise *wavelet* em sinais monofônicos captados em tempo real. Sons monofônicos são aqueles em que há apenas um tom presente em um intervalo de tempo definido, ao contrário dos sons polifônicos, que podem conter várias notas tocadas simultaneamente. A investigação de sons polifônicos, do ponto de vista computacional e musical, requer uma abordagem diferente, envolvendo a análise harmônica de acordes (MÜLLER, 2015).

Particularmente, a análise realizada objetivou associar os padrões sônicos aos símbolos de uma partitura musical a ser tocada, possibilitando a visualização dos eventos musicais classificados ao mesmo tempo em que são executados. O procedimento utilizado neste experimento seguiu as seguintes etapas:

- ETAPA 1: aquisição dos sinais sonoros de interesse por meio de um dispositivo de captação portátil para treinamento do sistema;
- ETAPA 2: aplicação da WPT ao sinal digitalizado, seguido da autocorrelação e da seleção do conteúdo espectral, a qual origina um vetor de características (VC);
- ETAPA 3: submissão do VC à rede neural para caracterização dos diversos eventos musicais;

Figura 35 – Sequência de etapas do método proposto.



Fonte: Elaborado pelo autor

- ETAPA 4: nova aquisição de sinais acústicos, captados em tempo real pelo dispositivo de captação portátil, objetivando avaliar o sistema;
- ETAPA 5: aplicação da WPT, da autocorrelação e da seleção do conteúdo espectral para formar os VCs, seguido do encaminhamento à rede neural para classificação;
- ETAPA 6: associação dos resultados da classificação aos símbolos partiturais.

A captação dos sinais acústicos foi realizada por um *smartphone*, entretanto, um *tablet*, ou outro gravador digital, também poderia ter sido utilizado, desde que as

especificações técnicas mínimas descritas na seção 4.1.3 fossem atendidas. O primeiro instrumento utilizado para produzir os sinais sonoros de interesse foi um teclado digital capaz de produzir sons tanto monofônicos quanto polifônicos. Para o método proposto, foi utilizado o mesmo tipo de timbre tanto na etapa de treinamento quanto na etapa de classificação, visto que timbres diferentes podem gerar conteúdos frequenciais distintos.

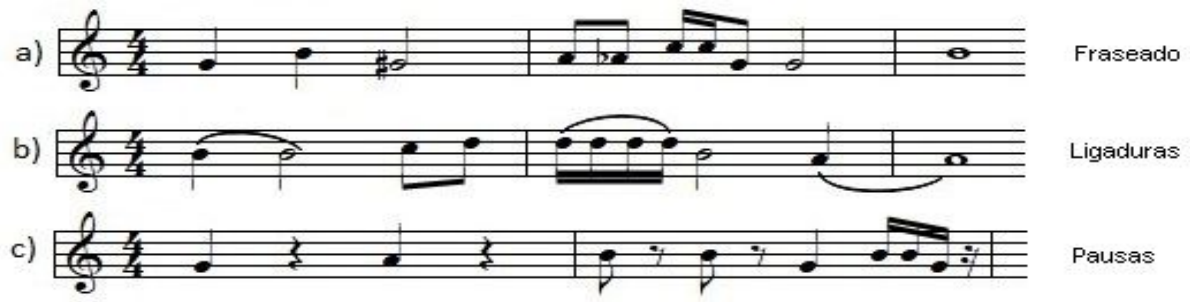
Para possibilitar também experimentos utilizando instrumentos não digitais, foi usado um saxofone alto. Normalmente, com exceção dos instrumentos eletrônicos, os geradores de sinais musicais são afinados usando alguma nota musical como referência. As situações nas quais determinado instrumento é afinado usando uma frequência ligeiramente diferente da esperada, caracterizando um instrumento “desafinado”, pode acarretar problemas na execução do método proposto, quando o treinamento da rede neural já tiver sido realizado. Isso porque todas as notas do instrumento são deslocadas em relação à frequência esperada, podendo-se tomar uma outra frequência no lugar da correta, ou às vezes nem mesmo classificá-la.

4.1.1 Seleção dos eventos musicais

Existe uma infinidade de eventos sonoros que podem ser aplicados a uma peça musical, desde uma simples sequência de notas musicais até fraseados mais complexos contedo modulações, expressões, trêmulos e outros efeitos. Entretando, para o processo de aprendizagem pelo qual um músico deve passar, uma classe de eventos básicos pode ser executada por diversos instrumentos. Neste trabalho, os eventos musicais a serem identificados são condizentes com o aprendizado inicial de um músico na prática e domínio básico de seu instrumento. Todos são símbolos pertencentes a partitura musical, ilustrados na Figura 36 e descritos a seguir:

- as notas musicais com sua entonação, de acordo com sua respectiva clave e posição dentro do pentagrama, e duração, de acordo com seu símbolo;
- as ligaduras sobre sons de mesma entonação, chamadas de ligadura de duração;
- as pausas, conhecidas como duração do silêncio entre os eventos;

Figura 36 – Eventos musicais a serem identificados pelo método proposto: (a) fraseados simples; (b) ligadura de duração e (c) pausas.



Fonte: Elaborado pelo autor

- o andamento, ou tempo de execução (bpm), da peça musical.

A maioria dos eventos amostrados compreendeu notas musicais situadas entre a segunda e sexta oitava do teclado digital, mais precisamente entre o Fá (F2) da segunda oitava e o Dó (C6) da sexta oitava. Na Figura 37 tem-se uma partitura na clave de sol com a extensão deste intervalo que compreende tal escala, sem os "acidentes". O Lá padrão (A4) do teclado com frequência em $440Hz$ corresponde à nota Fá sustenido (F3#) da terceira oitava do saxofone alto, considerando-se que a primeira oitava do saxofone se inicia na nota Lá sustenido mais grave (A1#).

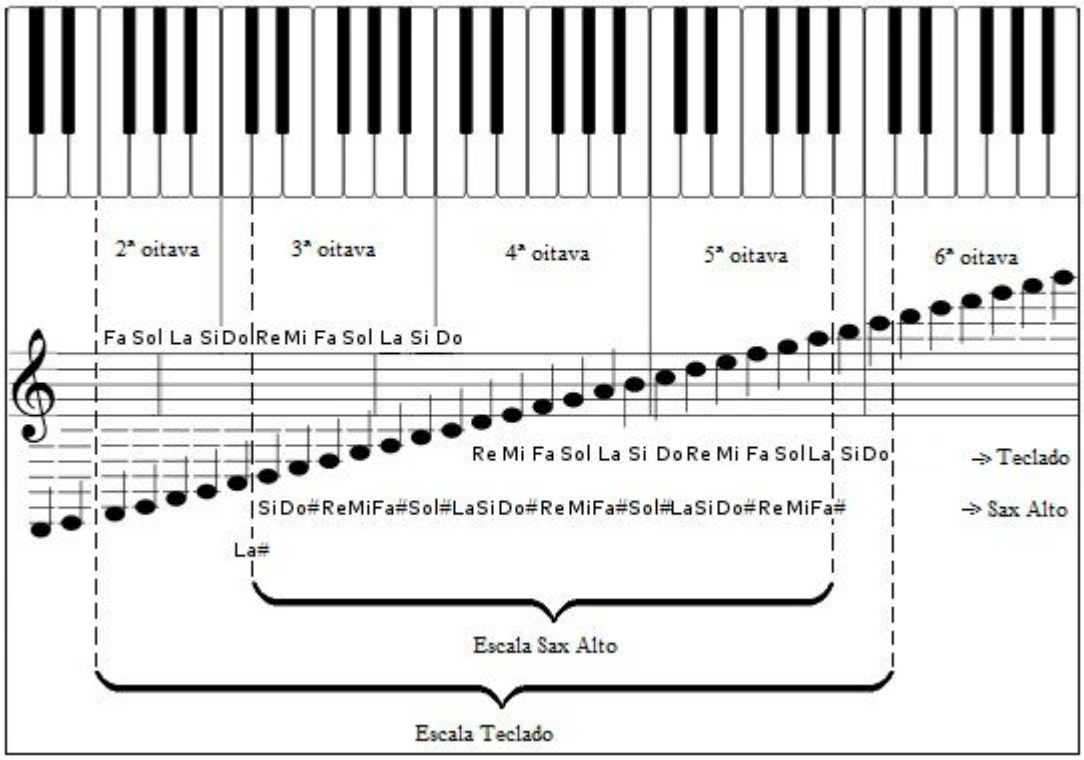
4.1.2 Amostragem do sinal

O processo de amostragem de sinais tem como objetivo digitalizar as formas de onda captadas pelo método proposto. Os parâmetros escolhidos para o processo de amostragem do modelo são uma resolução de 16 bits por amostra coletada a uma frequência amostral de $44100Hz$. Essa frequência foi escolhida em razão da capacidade de precisão de nossos ouvidos estarem na faixa máxima de $20000Hz$. Segundo o teorema da amostragem de Nyquist, é necessário o dobro da maior frequência para ser feita a discretização correta dos sinais.

4.1.3 Plataforma computacional e aquisição do sinal

O processo de treinamento da rede neural foi realizado separadamente, em uma máquina com 4GB de mémoria RAM e processador *dual-core* com velocidade de processamento igual a $2,2GHz$, utilizando o sistema operacional *Windows*. A rotina

Figura 37 – Notas musicais do teclado compreendidas e sua corresponde no saxofone alto



Fonte: Elaborado pelo autor

foi desenvolvida em linguagem Java, na IDE Eclipse. O treinamento diretamente no dispositivo móvel não foi possível em razão do grande número de exemplos requeridos para cada evento, resultando em um “estouro” de mémoria causado pela dimensão dos vetores (BIANCHI, 2014).

As rotinas de captação de áudio, decomposição *wavelet* e classificação foram implementadas para serem executadas em dispositivos móveis que utilizam o sistema operacional Android. O desenvolvimento de aplicações para tal plataforma é relativamente simples, podendo utilizar uma API Java e IDEs, isto é, ambiente de desenvolvimento, com licença livre. A ferramenta de desenvolvimento utilizada no trabalho foi a IDE Android Studio, disponibilizada pelo Google. Foi importada a biblioteca JWave, tratando-se de código aberto para implementação de *wavelets* ortogonais e bi-ortogonais.

A capacidade de processamento do método proposto é limitada aos recursos de *hardware* do dispositivo. A versão mínima do sistema operacional para execução do projeto é a Android 4.2, devendo haver capacidade de memória interna (RAM)

de no mínimo 1GB e processador com dois ou mais núcleos e velocidade igual ou superior a $1GHz$ de processamento. Essa configuração mínima é facilmente encontrada em *smartphones* e *tablets* comercializados atualmente. Neste trabalho, o acesso ao sinal de áudio capturado pelo microfone do dispositivo foi feito por meio da classe *AudioRecord*, que permite a configuração de diversos parâmetros para a leitura dos valores de entrada, como a taxa de amostragem desejada, a resolução das amostras de áudio (8 ou 16 bits) e o tamanho do *buffer* de amostras.

Os recursos de *hardware* do dispositivo estão diretamente ligados à classificação dos eventos de andamentos rítmicos com maior ou menor intensidade. Experimentos realizados utilizando a configuração mínima apresentaram uma média de cinco janelas temporais captadas por segundo, ou seja, cinco eventos musicais diferentes por segundo podem ser captados. Para um andamento de $60bpm$, considerando a semínima como unidade de tempo (uma semínima por segundo), a transição entre duas colcheias (duas colcheias equivalem a uma semínima) somam-se dois eventos diferentes, ou seja, um para cada colcheia, resultando em dois eventos distintos por segundo. Neste caso, a captação dos eventos ocorre sem problemas. Para um conjunto de quatro semicolcheias (quatro semicolcheias equivalem a uma semínima) tem-se um total de quatro eventos diferentes. As fusas e semifusas, equivalentes a $\frac{1}{8}$ e $\frac{1}{16}$ da semínima respectivamente, são impossibilitadas de captar no intervalo de um segundo. Neste caso, a solução é tocar a peça musical com um andamento mais lento, no caso uma semicolcheia a cada segundo. Músicos iniciantes geralmente tocam em um andamento mais lento quando se deparam com sequências de fusas e semifusas. Na Tabela 2, tem-se cada símbolo musical e a quantidade necessária de cada um para os andamentos de uma semínima por segundo e uma semicolcheia por segundo. Pode-se observar que o andamento de uma semicolcheia por segundo é suficiente para eventos envolvendo fusa e semifusas.

4.2 Decomposição *wavelet*

O sistema projetado utiliza um processo de decomposição *wavelet* para a representação dos sinais acústicos. Um bloco de 8192 amostras é submetido à WPT, gerando um conjunto de sub-sinais de aproximações e sinais de detalhes. Esse proce-

Tabela 2 – Quantidade de símbolos por segundo

Símbolo	Semínima por seg.	Semicolcheia por seg.
semibreve	$1/4$	$1/16$
mínima	$1/2$	$1/8$
semínima	1	$1/4$
colcheia	2	$1/2$
semicolcheia	4	1
fusa	N/A	2
semifusa	N/A	4

Fonte: Elaborado pelo autor

dimento resulta em uma projeção do sinal original em outro plano com dimensão de acordo com o nível de decomposição adotado. Os parâmetros relacionados ao nível de decomposição e a família de funções *wavelets* foram obtidos de forma prática, por meio de experimentos ao decorrer do estudo, conforme detalhado no próximo Capítulo.

A utilização da Transformada *Wavelet-Packet* no método proposto fornece uma resolução de tempo-frequência uniforme e igualmente distribuída para todas as sub-bandas de frequência. Em cada nível de decomposição, a transformação produz dois sinais de meia banda contendo as frequências baixas e altas do sinal de entrada que posteriormente são concatenadas para estabelecer o sinal transformado (OLIVEIRA, 2007a)(GUIDO, 2016)(SILVA; CARVALHO; MORET, 2006). Isso permite definir as sub-bandas correspondentes a oitavas musicais e notas musicais, considerando apenas aquelas necessárias para a análise de cada evento musical.

4.2.1 Família *wavelet* e nível de resolução

A utilização de *wavelets* diferentes leva a representações distintas do sinal analisado. Neste trabalho, um conjunto de famílias *wavelets* foi adotado para avaliar qual dentre elas apresenta a melhor acurácia para o objetivo proposto. Essas famílias são bem conhecidas e provaram ser vantajosas em inúmeras aplicações de processamento de sinais. Foram avaliadas as seguintes *wavelets*:

- Haar;
- Daubechies (DB2, DB4, DB8, DB12, DB16, DB20);
- Coiflets (C3,C5);

- Symlets (S8,S16);
- Beylkin (B9).

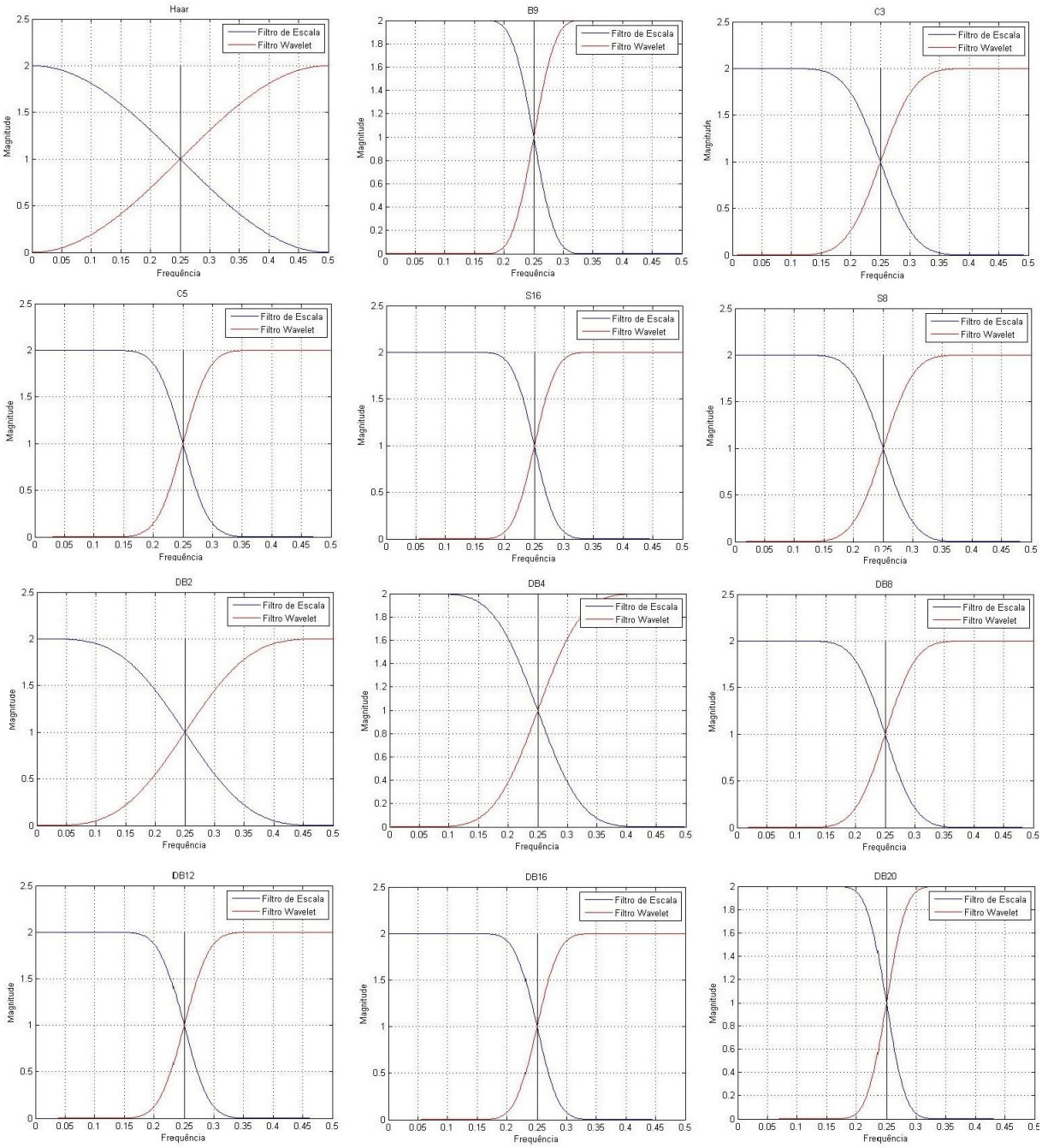
Para os filtros DBN, CN, SN e BN, onde N é o número de momentos nulos da função *wavelet*, verificaram-se diferenças de resposta em frequência. De acordo com os níveis de decomposição, as *wavelets* propiciaram uma melhor sintonia da janela de análise à medida que o número de momentos passou a ser maior. As respostas em frequência das doze funções de escala e *wavelet* utilizadas estão ilustradas na Figura 38.

Ao longo do desenvolvimento, notou-se que, quanto maior o número de momentos da função *wavelet*, mais a resposta em frequência real do filtro respectivo se aproxima da frequência ideal. A separação de bandas frequenciais permite que uma extensão de frequências específicas de interesse passe pelo filtro, enquanto que a extensão de frequências não desejadas seja atenuada. Filtros com bandas de transições mais estreitas permitem separar sinais de frequências próximas umas das outras. A atenuação da banda de corte deve ser alta o suficiente para eliminar frequências indesejadas. Estas características levam a um menor espalhamento de frequências, apresentando uma maior rejeição na banda não passante.

Percebeu-se durante as implementações que um maior número de coeficientes dos filtros passa-baixas e passa-altas empregado na decomposição *wavelet* acarreta um maior custo computacional. As bases DB20, C5, S16 e B9, apesar de apresentarem bons resultados quanto à localização frequencial, tiveram um custo computacional alto comparadas às bases com menos momentos. Pode-se observar que quanto menor a duração da função *wavelet* no tempo, mais as características de alta frequência são consideradas.

Um maior número de momentos mostrou-se melhor para representar sinais mais complexos. Isso se deve ao fato de a *wavelet* com um maior número de coeficientes ter uma resposta em frequência mais suave. Nas Figuras 39 e 40 têm-se dois sinais complexos representando a nota Dó (C3) captada pelo teclado digital e a nota Lá (A2) (relativa a C3 do teclado) pelo saxofone alto, com suas respectivas transformações utilizando a base DB2 e DB20. Em ambos os sinais, pode-se notar uma suavização no

Figura 38 – Respostas em frequência, passa-baixas e passa-altas, dos filtros *wavelet* utilizados

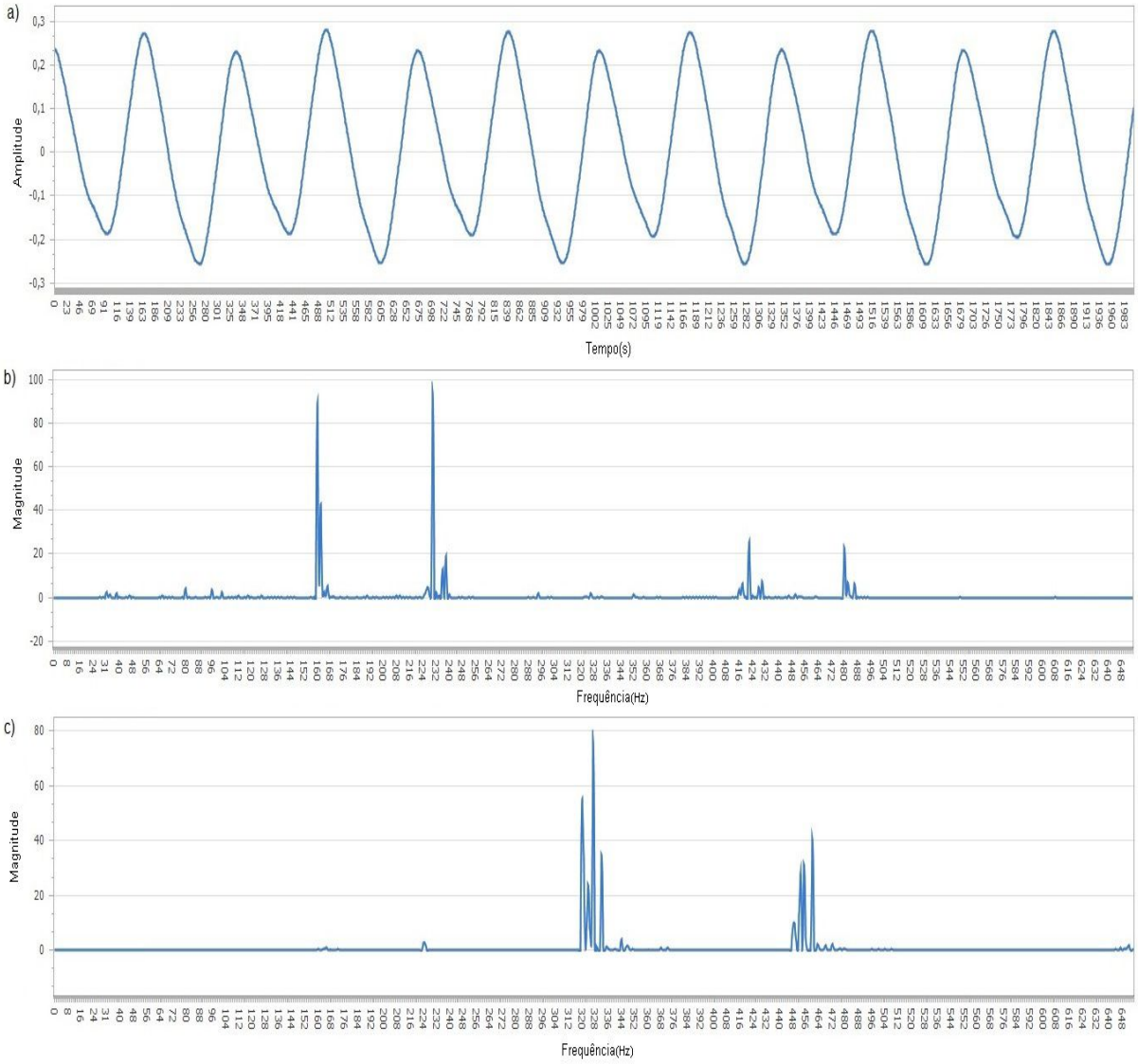


Fonte: Elaborado pelo autor

sinal transformado na DB20 em relação a DB2, uma redução do número de picos e um deslocamento de frequências resultando em uma melhor localização frequencial utilizando a Daubechies com maior número de momentos.

A separação de padrões sônicos em níveis de resoluções diferentes pode ser melhor verificada e destacada para algumas bases específicas. As bases de Symmlets

Figura 39 – Transformação *wavelet* do sinal da nota Dó (C3) no teclado digital. (a) sinal no domínio do tempo; (b) sinal transformado utilizando a DB2; (c) sinal transformado utilizando a DB20.

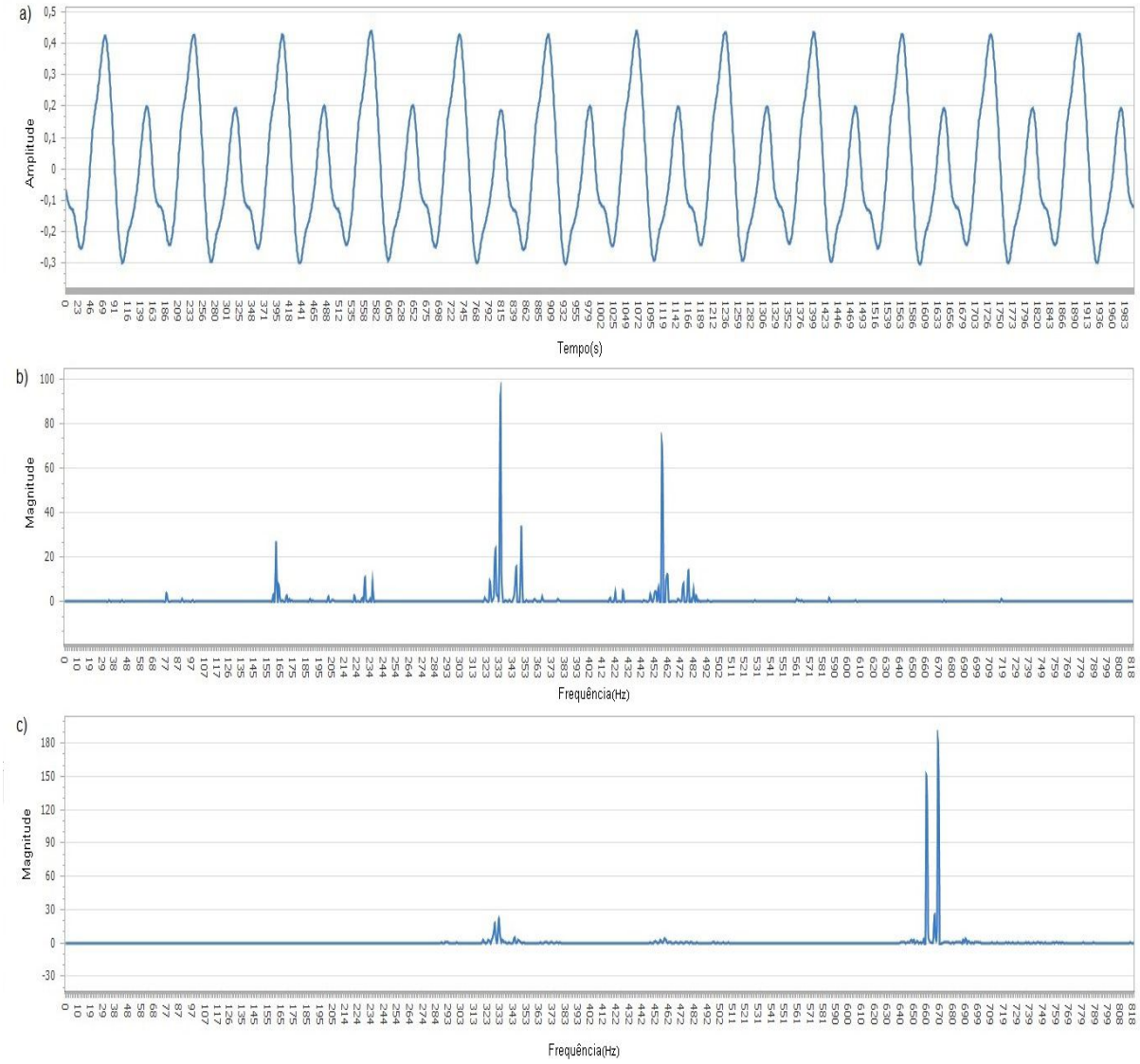


Fonte: Elaborado pelo autor

apresentaram melhores resultados de localização frequencial comparadas às bases de Coiflets, e um pouco similares às bases de Daubechies. Porém, tanto Symmlets quanto Coiflets apresentaram localização ruim para algumas faixas de frequências específicas. Notas musicais de baixa frequência próximas umas das outras como, por exemplo, um Dó (C3) e um Dó sustenido (C3#) da terceira oitava, apresentaram um mesmo índice frequencial no espectro.

A base de Beylkin (B9), conhecida na literatura como Beylkin 18, com 9 momentos nulos na função *wavelet* e 18 coeficientes, apesar de ser proposta para análise

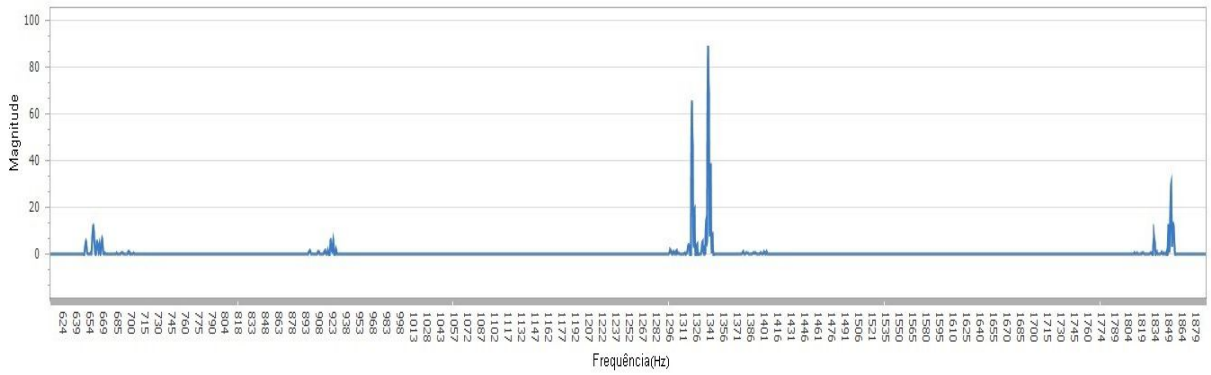
Figura 40 – Transformação *wavelet* do sinal da nota Lá (A2) no saxofone alto. (a) sinal no domínio do tempo; (b) sinal transformado utilizando a DB2; (c) sinal transformado utilizando a DB20.



Fonte: Elaborado pelo autor

de áudio em geral, não teve uma boa localização frequencial para o método proposto, comparada às outras famílias *wavelet* experimentadas. Na Figura 41 tem-se o mesmo sinal da Figura 40-(a), mas agora utilizando a base de Beylkin. Pode-se observar um grande deslocamento frequencial comparativamente a DB20, com um maior número de picos espalhados entre si. Eventos diferentes apresentaram padrões de conteúdos frequenciais semelhantes utilizando a B9, compromentendo assim a sua classificação.

Figura 41 – Sinal da Figura 40-(a) utilizando a base B9.



Fonte: Elaborado pelo autor

Os experimentos para escolha da base adotada procuraram obedecer critérios de custo computacional e melhor localização frequencial. As bases com menor quantidade de coeficientes se mostraram com menor custo computacional, porém mais sensíveis a ruídos, devido ao fato da transformação não poder comprimir a energia do sinal original em alguns valores de alta energia acima do limite de ruído. Os filtros com menor quantidade de coeficientes são geralmente empregados para a localização no tempo (ELFOULY et al., 2008).

A utilização da base de Haar mostrou baixa localização frequencial, mesmo nos níveis mais altos. É a base *wavelet* considerada mais simples, com apenas um momento nulo e suporte dos filtros com tamanho dois. Isso acarreta grandes mudanças não refletidas nos coeficientes de alta frequência. A *wavelet* de Haar geralmente é utilizada em sinais que possuem mudanças abruptas de valores no decorrer do tempo (ELFOULY et al., 2008). Os experimentos com essa base foram incentivados pelo baixo custo computacional, mostrando melhor desempenho de tempo de processamento em comparação com as outras bases. Porém, a base apresentou alta taxa de redundância em índices frequenciais de eventos distintos.

Dentre todas as bases experimentadas, a *wavelet* de Daubechies foi a que apresentou melhor localização frequencial para o método proposto. São *wavelets* ortonormais de suporte compacto, assimétricas, com ondulações contínuas, computacionalmente mais caras de usar do que a *wavelet* de Haar (ELFOULY et al., 2008). No entanto, a escala de sinais e as *wavelets* são ligeiramente mais longas, ou seja,

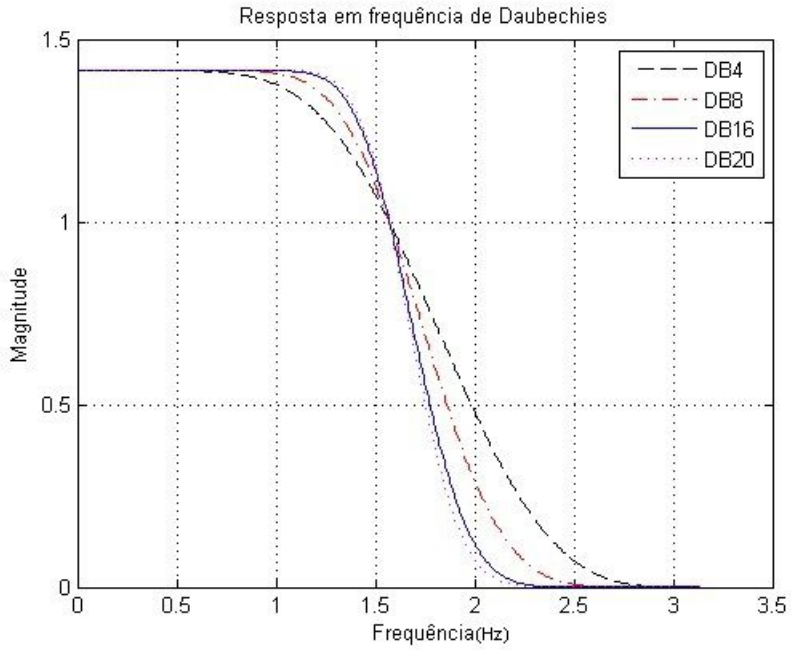
produzem médias e diferenças usando alguns valores a mais do sinal, provendo uma grande melhoria da capacidade dessas transformadas. As *wavelets* DB2, DB4, DB8, DB12, DB16 e DB20 foram testadas, mostrando um melhor localização frequencial para um maior número de coeficientes.

A cobertura da DB20 mostrou-se mais sintonizada, exibindo uma banda mais intensa e menor redundância de índices frequenciais para eventos diferentes, comparada às bases de Daubechies com menor número de coeficientes. Isso resultou em uma cobertura de faixas frequenciais mais localizadas, isolando mais alguns componentes em níveis diferentes. Apesar das vantagens citadas, a DB20 mostrou um alto custo computacional. Em geral, as *wavelets* de ordem superior possuem melhor resposta em frequência do que as de menor ordem, no entanto, o custo computacional aumenta conforme a ordem da *wavelet* aumenta.

As respostas em frequência dos filtros *wavelet* de Daubechies são mostrada na Figura 42. Pode-se observar que à medida que seu comprimento aumenta, mais pontos são próximos de zero. Para sua escolha, deve existir uma compensação entre uma boa resposta de frequência de filtros mais longos e a maior complexidade exigida. Uma resposta em frequência mais plana na sua banda de passagem e banda de corte evita flutuações impróprias quando o sinal apresenta um ganho reduzido ou excessivo em algumas sub-bandas causado por magnitudes de frequências imprecisas.

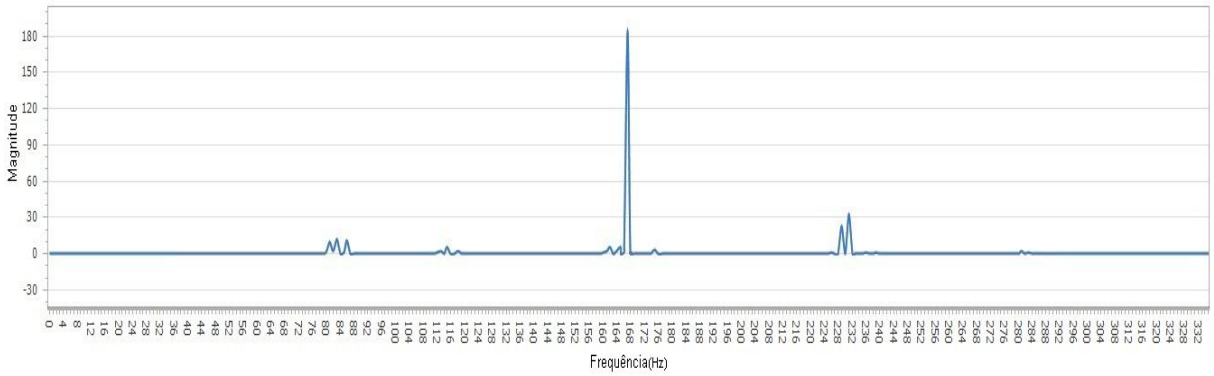
Com um compromisso entre os requisitos de computação do tempo e a taxa de reconhecimento correta, a base DB16 foi considerada a melhor opção de *wavelet* para o método proposto, pois apresentou uma taxa de reconhecimento similar a DB20 e um custo computacional mais baixo, tornando-se interessante para a aplicação em tempo real. Sua cobertura mostrou-se mais localizada, em razão de sua banda passante ser mais intensa e sua banda de corte ser bem mais atenuada em relação às Daubechies de ordem menor. A DB16 destacou-se por apresentar menor redundância na representação frequencial para eventos diferentes nas baixas frequências, mantendo uma boa localização para frequências mais altas. A maior suavidade no decaimento das bandas de transição e corte levou a uma melhor localização em frequências, resultando em uma melhor separação dos padrões musicais. O método proposto utilizando a autocorrelação apresentado na seção 4.3 reduziu os picos correspondentes às harmônicas,

Figura 42 – Resposta em frequência de DB4, DB8, DB16 e DB20.



Fonte: Elaborado pelo autor

Figura 43 – Sinal da Figura 40(a) utilizando a base DB16



Fonte: Elaborado pelo autor

permitindo a base DB16 apresentar um bom compromisso entre detecção frequencial e eficiência computacional.

A construção da Daubechies DB16 resulta em uma coleção de 32 coeficientes (NIEVERGELT, 1999) de escala. Na Figura 43 tem-se o mesmo sinal da Figura 40-(a), mas agora utilizando a base DB16.

Desconsiderando a informação temporal, a melhor resolução espectral é garantida sempre que o nível de decomposição mais profundo é adotado (GUIDO, 2017). O nível m de resolução escolhido foi o nível máximo permitido pelo processo de

decomposição. Ele foi determinado pela equação 4.1.

$$m = \frac{\log(N)}{\log(2)} \quad , \tag{4.1}$$

sendo N o número total de amostras. Assim, para janelas de tamanho $N = 8192$, o nível máximo é $m = 13$. Níveis mais baixos de resolução mostraram um custo computacional melhor, porém para o método proposto, a localização frequencial ficou comprometida, apresentando redundâncias de picos máximos tanto para baixas frequências quanto para altas frequências. O nível máximo de resolução permitiu uma melhor classificação dos eventos, com uma boa eficiência computacional.

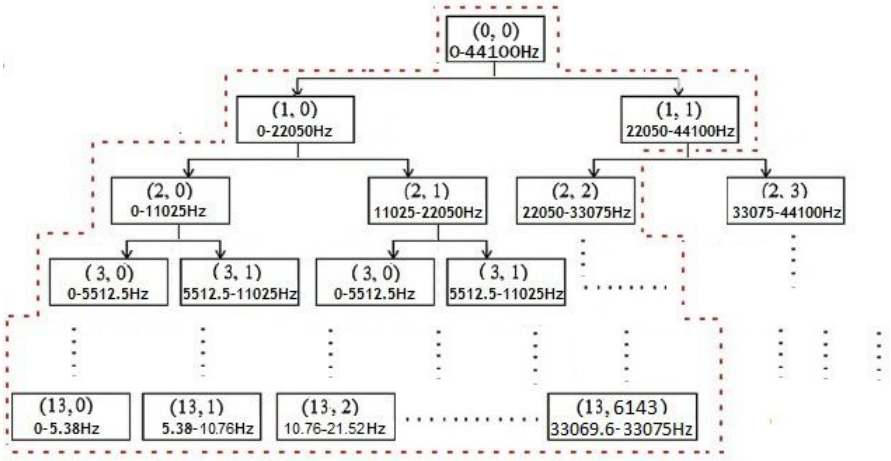
Uma descrição refinada de tempo-frequência com base nas estratégias previamente definidas requer o uso de um operador de energia elementar para converter todas as amostras de cada sub-banda em seus potenciais instantâneos. Dado um sinal x transformado, para a amostra n de cada sub-banda $b = 0, 1, ..., 2^j - 1$ do nível mais profundo j, tem-se:

$$x_{n,b} = (x_{n,b})^2. \tag{4.2}$$

O tempo de processamento máximo necessário para a análise em tempo real mostrou-se razoavel quando um conjunto de 8192 amostras serviu como entrada à WPT, em comparação com o número de amostras anterior. Porém, esse número mostrou-se insuficiente para uma boa localização frequencial: um total de 13 níveis de decomposição *wavelet* funciona bem para altas frequências, mas apresenta um maior número de redundâncias de índices para eventos diferentes em detecções de frequências mais baixas. A decomposição em árvore utilizada, com as respectivas faixas de frequência para cada nível, está ilustrada na Figura 44.

Em cada nível decomposto, a WPT transforma o sinal em um novo sinal com o mesmo número de amostras, servindo como entrada para o nível subsequente. A decomposição do sinal foi realizada tanto no sinal de aproximação, quanto no sinal de detalhes, como visto na seção 3.7. A Tabela 3 contém um resumo de cada faixa espectral obtida e sua respectiva resolução por meio dos níveis. A resolução de $2,69Hz$ para cada sub-banda foi suficiente para uma análise frequencial mais refinada de cada evento musical.

Figura 44 – Decomposição *wavelet* em 13 níveis.



Fonte: Elaborado pelo autor

Tabela 3 – Faixas frequenciais decompostas em cada nível e suas respectivas resoluções

Nível	Faixa	Resolução
1	0 - 22050Hz	11025Hz
2	0 - 11025Hz	5512,5Hz
3	0 - 5512,5Hz	2756,25Hz
4	0 - 2756,25Hz	1378,125Hz
5	0 - 1378,125Hz	689,06Hz
6	0 - 689,06Hz	344,53Hz
7	0 - 344,53Hz	172,265Hz
8	0 - 172,265Hz	86,13Hz
9	0 - 86,13Hz	43,065Hz
10	0 - 43,065Hz	21,53Hz
11	0 - 21,53Hz	10,765Hz
12	0 - 10,765Hz	5,38Hz
13	0 - 5,38Hz	2,69Hz

Fonte: Elaborado pelo autor

O deslocamento de uma ou mais oitavas tanto para cima como para baixo resulta diretamente no deslocamento de padrões sônicos, alterando a identificação e a percepção dos componentes e estruturas musicais. Isso porque a concentração de suas características está relacionada diretamente ao seu conteúdo frequencial. As notas mais graves, por apresentarem frequências mais baixas, se concentram nos níveis mais baixos de resolução do que as notas mais agudas, por apresentarem frequências mais altas. Eventos musicais apresentando uma mesma expressividade e dinâmica possuem descrições semelhantes em níveis consecutivos, porém para

eventos tocados em oitavas diferentes, o conteúdo frequencial não ocupou os mesmos níveis. A mudança de oitava em uma mesma nota, apesar de apresentar harmônicos semelhantes, mostrou diferentes deslocamentos de padrões frequenciais.

4.3 Aplicação da autocorrelação

O número abundante de componentes harmônicos contidos em sinais sonoros é um fator que pode comprometer a acurácia na detecção de eventos. Alguns problemas importantes são encontrados, como a variação da frequencia fundamental no tempo ou o aparecimento de sub-harmônicos que pode causar falsa detecção. A aplicação da autocorrelação pode ser utilizada para encontrar padrões de repetição em um sinal contendo vários componentes harmônicos. Vários métodos já foram propostos utilizando a autocorrelação, com o objetivo de reduzir o efeito dos múltiplos harmônicos, como a extração de batida temporal para estimação de tempo musical, proposto por Hu (HU, 2010), e a classificação de gêneros musicais descrito por Popescu et al. (POPESCU; GAVAT; DATCU, 2009).

A decomposição *wavelet* aplicada sobre o sinal de entrada resulta em vários componentes harmônicos, além da presença ou não de uma frequência fundamental f_0. Por essa razão, a detecção do maior pico encontrado no sinal transformado revela, em alguns casos, valores de harmônicos ao invés da f_0. Esse resultado tende a comprometer a classificação do sinal musical quanto à sua frequência, apresentando índices de picos máximos encontrados semelhantes em diferentes notas musicais tocadas. Em outras palavras, uma frequência harmônica é tomada nesse caso como uma frequência fundamental.

Foi adotada, então, uma estratégia para reduzir os picos correspondentes às harmônicas de uma dada frequência f_0. A detecção de periodicidade baseia-se em uma "autocorrelação generalizada", dada por:

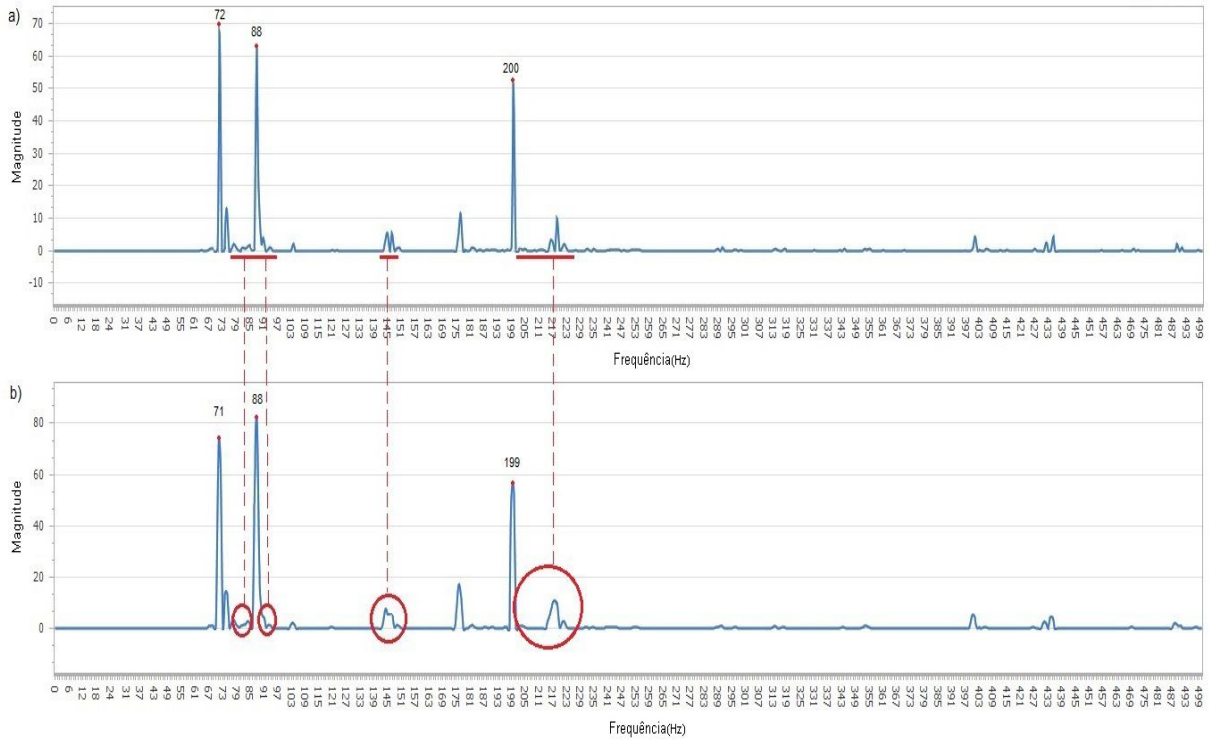
$$y(n) = WPT(x(n))^2 + WPT(x(n+1))^2 \qquad , \tag{4.3}$$

onde WPT representa a Transformada *Wavelet-Packet* aplicada sobre o sinal $x(n)$, e n representa o índice temporal. O sinal resultante da aplicação da WPT é submetido à

autocorrelação originando um novo sinal $y(n)$ com a mesma dimensão de $x(n)$, mas agora com o número de picos reduzidos em comparação à $WPT(x(n))$.

Na Figura 45, observa-se um exemplo da aplicação da autocorrelação sobre o sinal sonoro da nota Si (B2) da segunda oitava do saxofone alto. Na Figura 45-(a), tem-se o sinal resultante da aplicação da WPT, seguida pela autocorrelação aplicada sobre ele, representado pela Figura 45-(b). É possível notar que o processo de autocorrelação resultou numa suavização nas áreas circuladas, reduzindo o número de máximos locais e consequentemente os harmônicos. Houve também um deslocamento dos maiores índices frequenciais: em (a) os três maiores picos correspondem aos índices 72, 88 e 200; em (b) os três maiores picos correspondem aos índices 71, 88 e 199. Esse deslocamento reduziu o número de redundâncias encontradas entre dois sinais de frequências diferentes, em razão de um mesmo harmônico estar presente em ambos os sinais.

Figura 45 – Autocorrelação da nota Si (B2) da segunda oitava do saxofone alto: (a) sinal originado diretamente da WPT e (b) sinal resultante da autocorrelação.



Fonte: Elaborado pelo autor

4.4 Seleção de picos candidatos

Os picos máximos resultantes da função de autocorrelação correspondem às periodicidades encontradas no sinal. Uma vez calculada, a função de autocorrelação reduz o efeito dos múltiplos inteiros das periodicidades, tomando um pico com índice de tempo n, e eliminando ou atenuando o possível pico com índice de tempo igual a $n + 1$. Para melhorar a detecção dos picos máximos, a função de autocorrelação pondera mais fortemente os picos que têm amplitudes maiores do que aqueles com amplitudes menores, como ilustrado na Figura 45.

O maior pico encontrado geralmente corresponde ao primeiro pico, implicando a frequência fundamental. Porém, podem existir picos maiores que o primeiro, por causa de harmônicos que não foram eliminados na autocorrelação. As técnicas de limiarização de coeficientes de uma série *wavelet* têm como objetivo a redução, ou mesmo eliminação, de valores indesejados em um sinal.

O procedimento consiste em eliminar os picos para os quais as amplitudes estiverem abaixo de um certo limiar relativo ao pico de maior amplitude. Neste trabalho, é usado o limiar universal proposto por Donoho e Johnstone (DONOHO; JOHNSTONE, 1994), calculado da seguinte forma:

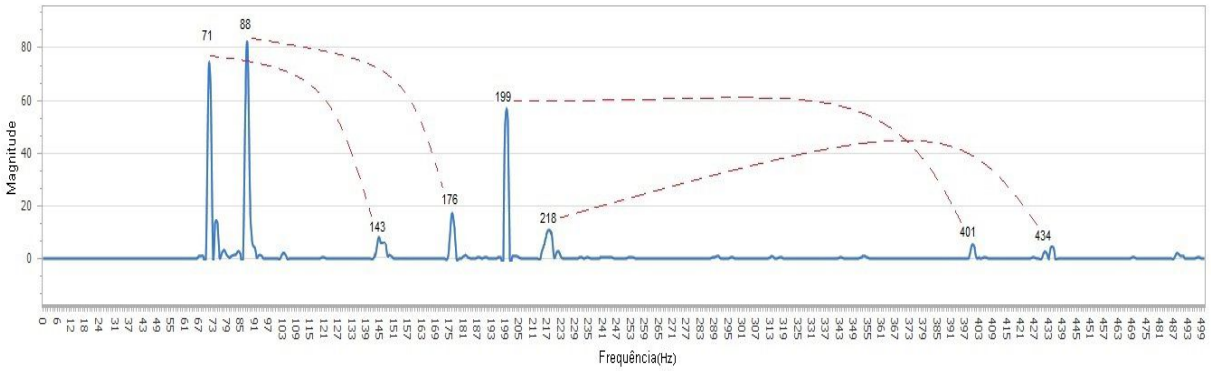
$$\lambda = \sigma\sqrt{2\log_{10}(N)} \qquad , \tag{4.4}$$

onde N corresponde ao comprimento do sinal e σ representa a estimativa de ruído, dado por: $\sigma = \frac{mediana(|C_A|)}{0{,}6745}$, sendo C_A os coeficientes wavelet de aproximação. Uma alteração foi realizada, onde a mediana dos coeficientes de aproximação é substituida pela média dos coeficientes wavelet-packet C_{wp} calculados, resultando em $\sigma = \frac{media(|C_{wp}|)}{0{,}6745}$. Dois índices são escolhidos representando o primeiro maior pico e o primeiro pico cujo valor é maior que o limiar encontrado e que ainda não tenha sido escolhido. Os índices são ordenados em ordem descrescente, para evitar padrões diferentes com os mesmos índices, e servem então como um padrão de conteúdo frequencial do evento analisado.

Os picos máximos encontrados em função da aplicação da autocorrelação e dos limiares são bons indicativos de frequências fundamentais ou séries harmônicas providas de fundamentais presentes no sinal. Na Figura 46 tem-se o mesmo exemplo da Figura 45-(b), mas agora com indicativos de relação entre os harmônicos presentes.

A frequência fundamental é caracterizada como o primeiro harmônico, provendo a partir dela componentes múltiplos como demais harmônicos. O conteúdo harmônico presente é caracterizado pelos pares de componentes com índices iguais ou próximos de seu múltiplo. Esses conteúdos se apresentam diferentemente para distintos eventos musicais, possibilitando a sua classificação pelos diferentes padrões apresentados.

Figura 46 – Conteúdo harmônico encontrado após a autocorrelação



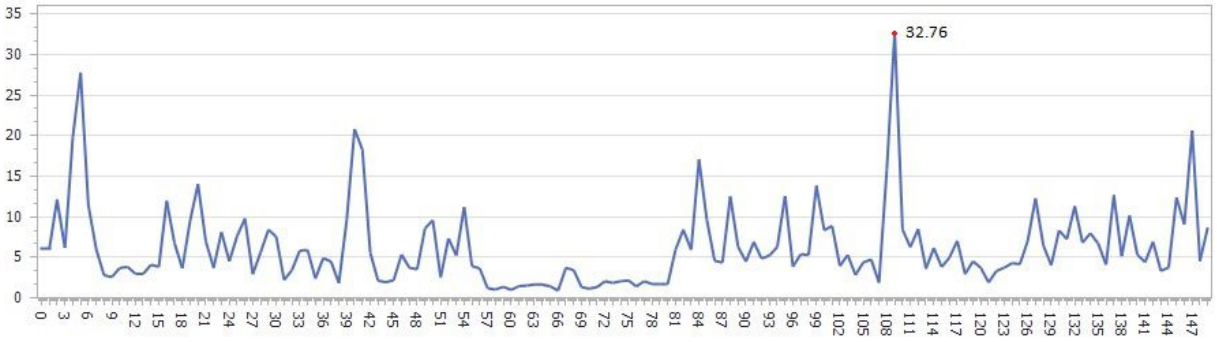
Fonte: Elaborado pelo autor

4.5 Eventos de pausas e duração

Os eventos de pausa, ou seja, silêncio, são aqueles em que não há uma ocorrência de som durante a execução musical. Analisando os valores de limiar calculados durante o processo de seleção de picos, nota-se que os valores mantém um padrão baixo em comparação aos limiares de eventos sonoros. O menor valor de limiar encontrado para eventos sonoros foi de 1000. Na figura 47 têm-se os valores de limiar capturados no período de silêncio correspondente à 150 amostras, onde o maior valor capturado corresponde a $32,76$. Sendo assim, são considerados zonas de silêncio todos os eventos com valores de limiar menores que 1000.

A duração de cada evento musical é validada no processo de captura, calculando-se o tempo em que um mesmo evento é identificado no decorrer da captura de cada janela temporal, até ocorrer uma nova identificação de um evento diferente. Por exemplo, ao soar a nota Lá(A), calcula-se por quanto tempo manteve-se o seu som, até ocorrer uma mudança para um evento diferente. Deste modo, é possível classificar cada figura musical de acordo com sua respectiva duração (mínima, semínima, colcheia

Figura 47 – Valores de limiar capturados correspondentes a 150 amostras



Fonte: Elaborado pelo autor

etc (PRIOLLI, 2015)) e identificar a ocorrência de ligaduras de valor (eventos de mesma altura ligados).

4.6 A Rede Neural RBF

Para que o método proposto possa reconhecer e classificar os eventos musicais analisados, foi adotada uma RNA como mecanismo de reconhecimento e classificação. Optou-se por utilizar uma rede do tipo RBF, em razão da possibilidade da camada oculta ser configurada com um número de neurônios igual ao número de exemplos de treinamento e, ainda, em função da dispersão particular exibida pelos dados de cada classe. Além disso, a simplicidade e eficiência computacional, utilizando o paradigma da aprendizagem híbrida, tornaram-se um ponto forte para sua escolha. Na primeira etapa, com um aprendizado não-supervisionado, foram necessários apenas alguns ajustes nos núcleos gaussianos. Na segunda etapa, supervisionada, tem-se um ajuste exato pela solução de um sistema linear possível e determinado que conduz aos pesos existentes entre a camada oculta e a camada de saída.

Para o treinamento da rede, vários testes para diferentes configurações quanto ao número de elementos de entrada e elementos ocultos da rede foram utilizados, mantendo-se sempre apenas um elemento na camada de saída, ou seja, um neurônio para classificar o evento musical. O resultado do neurônio de saída é relacionado com a frequência do evento musical analisado.

Partiu-se inicialmente com uma configuração onde a camada de entrada foi

composta com um número de amostras idêntico ao número de componentes resultantes da WPT, ou seja, 8192 elementos. Em termos de eficiência computacional, esse modelo de configuração não se mostrou capaz de realizar a classificação dos eventos em tempo real. O número de neurônios da camada oculta é calculado obtendo-se o número de exemplos de treinamento total. Foram necessários previamente, para uma classificação correta, 50 exemplos para um único evento x, resultando em uma camada oculta de $50 \cdot x$ elementos. O custo computacional aumentou quando foi necessário ampliar o número de exemplos de acordo com o número possível de eventos musicais a serem classificados. Quando alterados os valores de amplitude (volume do som) do sinal, foram necessários novos exemplos de treinamento, aumentando o número de neurônios da camada oculta e resultando assim em um maior custo computacional.

Devido à alta taxa de valores aleatórios dos coeficientes resultantes da aplicação da WPT, optou-se por configurar a camada de entrada da RNA de acordo com o número de padrões frequenciais capturados. Isso porque até mesmo quando a intensidade de som é alterada, novos exemplos de treinamento são necessários. A classificação pelos índices frequenciais diminuiu o número de exemplos de treinamento necessários, visto que a localização das maiores concentrações de energia no domínio frequencial varia pouco para um mesmo evento musical capturado por um mesmo instrumento. A aplicação da autocorrelação como visto na seção 4.3, reduziu o número de redundâncias para eventos diferentes, contribuindo diretamente para a escolha dessa configuração.

4.6.1 Arquitetura da rede RBF utilizada

A rede foi implementada com dois neurônios na camada de entrada, recebendo cada VC composto pelas posições dos picos fornecidos pelo processo de seleção descrito na seção 4.4, encontrados nas 8192 amostras resultantes da decomposição e autocorrelação. A camada oculta contém um número total de neurônios igual ao número total de exemplos de treinamento fornecidos de acordo com o instrumento musical escolhido. O objetivo da camada oculta é separar os padrões de entrada não linearmente separáveis em um conjunto de saídas linearmente separáveis. Para esta camada utilizou-se uma função do tipo Gaussiana como função de ativação. A camada

de saída foi implementada com um neurônio, representando o evento musical a ser classificado, com um rótulo numérico particular para caracterizá-lo.

O número de exemplos de treinamento foi obtido por meio de experimentos, quando foi observado que um número maior que 500 apresentou melhor classificação, porém com um custo computacional maior. Um número menor do que 50 mostrou-se não suficiente para o cálculo dos pesos sinápticos, uma vez que o número de padrões de picos encontrados capturados por evento foi alto, em torno de 250 padrões diferentes. Adotou-se, então, o número de exemplos igual ao maior número de padrões encontrados para uma classe de eventos. Os eventos dessa classe que apresentam uma quantidade de padrões menores são completados com a repetição dos primeiros exemplos, distintos apenas em termos de normalizações.

4.6.2 Treinamento e reconhecimento

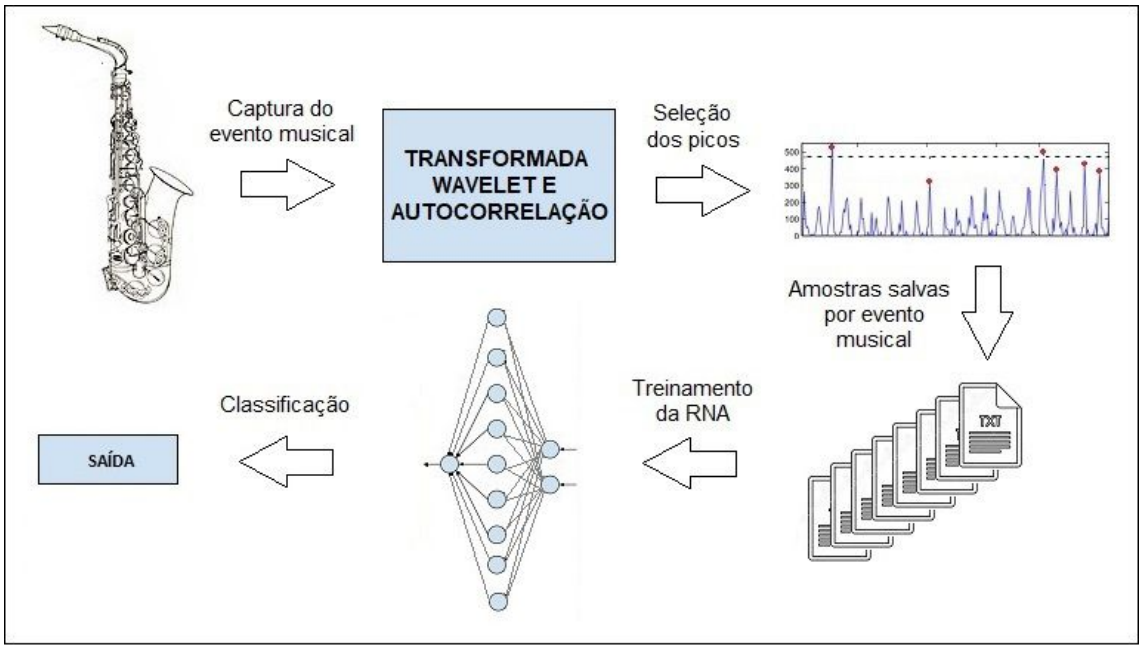
O processo de aprendizagem iniciou-se pela captura dos dois picos por evento musical para um dado instrumento. São colhidos 150 exemplos de cada evento para o teclado digital e 250 exemplos para cada evento do saxofone alto, gravados em um arquivo de texto representando cada evento musical independente. Cada instrumento tem seus arquivos referentes a cada evento musical que pode ser executado. O processo é ilustrado na Figura 48. Depois de salvos, os arquivos podem ser submetidos à RNA para o treinamento.

O comportamento do sinal resultante da decomposição e autocorrelação permite identificar diferentes comportamentos das harmônicas por meio dos picos máximos resultantes. Na Figura 49, vemos o espectro harmônico de três diferentes eventos concedidos pelo saxofone alto. Pode-se observar que cada evento musical tem um comportamento espectral distinto, ou seja, um padrão harmônico que, espera-se, a rede consiga identificar após ter sido treinada. Os arquivos contendo os padrões harmônicos capturados são utilizados como um vetor de entrada para treinar a rede. Por exemplo, para o seguinte entrada:

$$\text{Entrada} = [244, 212] \tag{4.5}$$

representando a nota Lá (A2) da segunda oitava do saxofone alto, a saída desejada

Figura 48 – Processo de treinamento da RNA.



Fonte: Elaborado pelo autor

para esse instrumento é definida como:

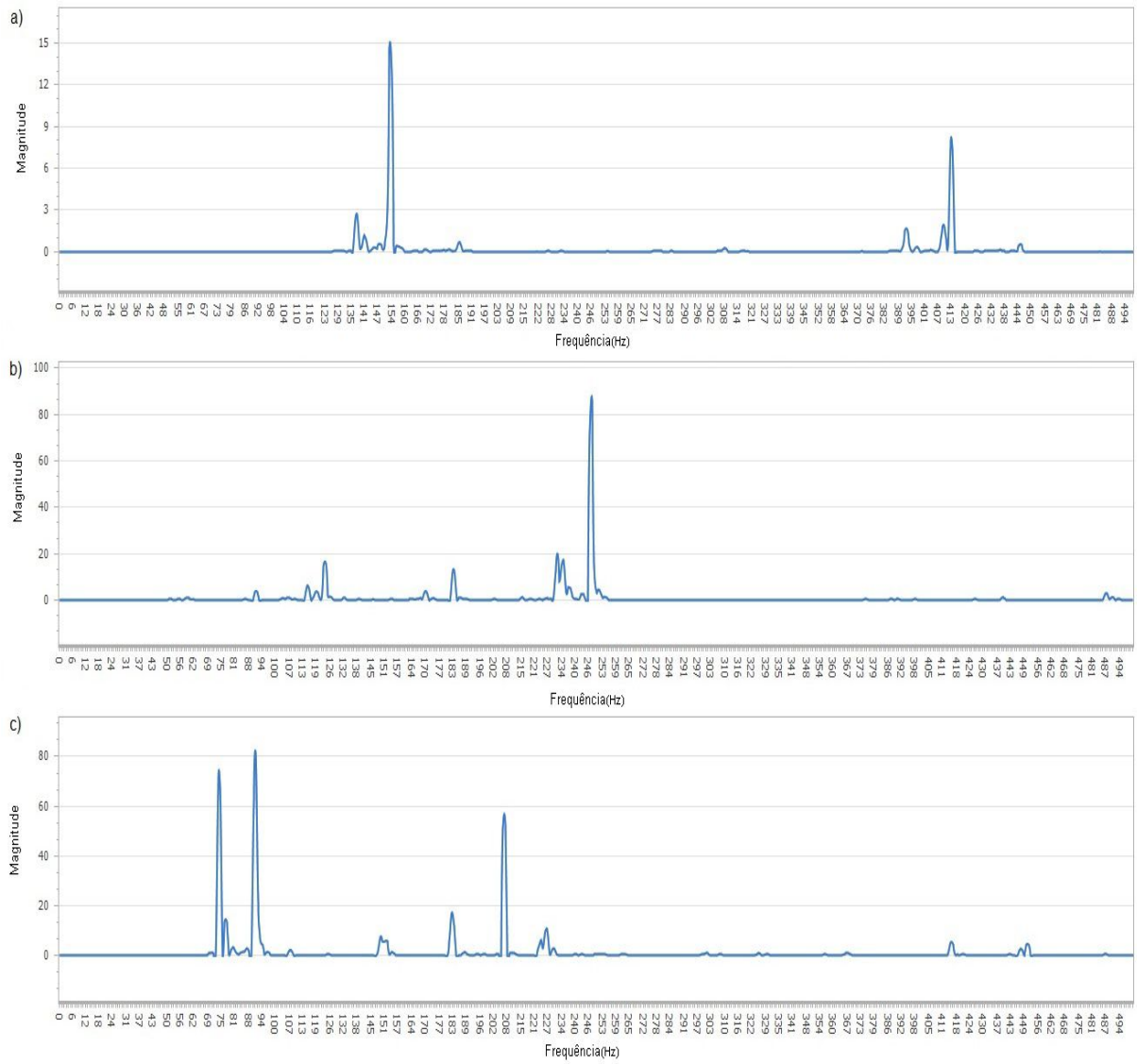
$$\text{Saída} = [440] \tag{4.6}$$

onde o número 440 representa a frequência referente à nota tocada. O treinamento da rede foi feito por instrumento, ou seja, cada instrumento musical deve ter um vetor de pesos sinápticos independentes. Isso porque a rede não foi capaz de identificar os mesmos eventos musicais para instrumentos musicais diferentes. O comportamento dos padrões harmônicos apresentou-se diferente para os mesmos eventos, e em alguns casos, igual para eventos diferentes, compromentendo a classificação. Foi necessário, então, treinar a rede independentemente para cada instrumento musical. Após o treinamento, com os vetores de pesos sinápticos com seus respectivos valores devidamente calculados, a RNA pôde ser submetida à execução em tempo real.

4.7 Execução em tempo real

Após o treinamento da RNA com as amostras dos eventos musicais armazenadas em arquivo, foi possível executar o sistema em tempo real, ou seja, reproduzir acusticamente um evento musical por um determinado instrumento, e o sistema ime-

Figura 49 – Três eventos diferentes concedidos pelo saxofone alto: (a) nota Dó (C3); (b) nota Sol (G2) e (c) nota Si (B2).



Fonte: Elaborado pelo autor

diatamente informar qual evento foi reconhecido. Os exemplos de sinais de áudio submetidos devem ser do mesmo instrumento no qual a RNA foi treinada.

5 TESTES E RESULTADOS

Para avaliar a influência do tamanho da janela temporal e dos filtros *wavelet*, os seguintes experimentos foram realizados:

- Experimento 1: Com o objetivo de escolher uma dimensão de janela temporal, mantendo-se o compromisso entre boa resolução frequencial e baixo custo computacional, fixou-se uma função *wavelet* e variou-se o comprimento da janela. Para este experimento, foram colhidos exemplos de sinais referentes a cada evento musical de interesse, no caso as notas musicais. A função inicialmente escolhida foi a *wavelet* de Daubechies com 20 momentos nulos (DB20), devido a sua resposta em frequência.
- Experimento 2: Para avaliar a influência dos filtros, um conjunto de famílias *wavelet* foi adotado para avaliar qual dentre elas apresentava uma melhor acurácia para o método proposto. Fixou-se um número de exemplos de casos, de acordo com o experimento 1, e variou-se cada filtro *wavelet* e o tamanho do seu suporte.

As etapas do experimento foram divididas em duas fases: treinamento e testes. O treinamento consiste em captar e gravar blocos de sinais contendo cada evento musical de seu respectivo instrumento e submetê-los ao método proposto. Em seguida, os vetores resultantes de cada evento foram selecionados para treinar a rede neural RBF. Após o treinamento, o sistema foi submetido à execução em tempo real.

Nos experimentos, a acurácia foi avaliada por meio da taxa de acertos, calculada da seguinte forma:

$$Taxa = \frac{100 \cdot a}{j}\% \tag{5.1}$$

onde a é o número de janelas corretamente classificadas e j o total de janelas. Inicialmente optou-se por um número de 250 janelas por evento antes da escolha do tamanho de janela e do filtro *wavelet* ideal. Para o experimento 1, foram testadas as janelas com comprimento de 4096, 8192, 16384, 32768 e 65536 amostras. Para o experimento 2, foram testadas as *wavelets* de Haar, Symlet, Coiflet, Beylkin e Daubechies, variando-se

o tamanho de seu suporte. Em ambos os experimentos, foi utilizado sempre o nível máximo de resolução.

Nas Tabelas 4 e 5, apresentam-se os resultados dos experimentos 1 e 2, realizados no saxofone alto. Nas Tabelas 6 e 7, têm-se os resultados dos experimentos 1 e 2 realizados no teclado digital. A taxa de acertos refere-se às frequências de determinadas faixas de eventos musicais do saxofone alto e do teclado digital.

Tabela 4 – Resultados do experimento 1 - Saxofone alto

Tamanho	A#1 a D#2	E2 a G#2	A2 a C#3	D3 a F#3	G3 a C#4	D4 a F#4	J\s
4096	50,20	53,52	71,33	71,66	70,0	67,30	11
8192	72,80	75,35	90,0	92,58	90,30	89,20	6
16384	73,65	77,77	91,33	94,0	90,0	91,55	2
32768	80,0	81,44	93,26	95,48	95,32	93,66	1
65536	86,12	89,85	98,55	98,20	97,62	94,92	0,5

Fonte: Elaborado pelo autor

Tabela 5 – Resultados do experimento 2 - Saxofone alto

Wavelet	A#1 a D#2	E2 a G#2	A2 a C#3	D3 a F#3	G3 a B3	C4 a F4
haar	36,58	34,10	49,56	51,63	50,52	48,20
sym8	70,10	65,50	74,25	74,30	73,28	70,50
sym16	72,50	69,64	70,26	74,26	71,0	76,50
coif3	52,80	54,36	52,35	53,65	62,30	60,60
coif5	54,60	5745	55,55	64,78	65,50	62,76
beyl9	45,0	38,30	39,65	55,0	58,96	56,68
daub2	44,45	42,0	53,80	55,50	54,0	48,50
daub4	64,68	63,32	65,86	66,70	65,50	63,40
daub8	72,34	72,50	76,50	77,75	77,05	73,0
daub12	74,10	73,80	76,45	78,56	77,05	74,65
daub16	80,38	77,84	89,40	92,93	91,36	88,39
daub20	72,80	75,35	90,00	92,58	90,30	89,20

Fonte: Elaborado pelo autor

Observando-se os resultados das Tabelas 4 e 6, nota-se que as janelas com dimensão de 8192 amostras proporcionaram uma taxa de acertos mais alta, considerando-se o compromisso entre resolução frequencial e baixo custo computacional. Apesar das janelas com maior comprimento apresentarem melhores resultados relativos ao aumento de seu suporte, o número de janelas por segundo (J\s.) apresentou-se baixo, comprometendo o objetivo de identificação em tempo real. Por meio dos resultados obtidos nas Tabelas 5 e 7, a base DB16 foi considerada a melhor opção de *wavelet*

para o método proposto, pois apresentou uma menor redundância na representação frequencial para eventos distintos. A base DB20 apresentou resultados ligeiramente próximos da DB16, porém com um custo computacional mais alto.

Tabela 6 – Resultados do experimento 1 - Teclado digital

Tamanho	F2 a B2	C3 a F#3	G3 a C#4	D4 a G#4	A4 a E5	F5 a C6	J\s
4096	72,0	68,36	69,44	73,22	65,63	72,05	11
8192	92,30	91,88	91,50	92,33	88,25	92,90	6
16384	94,88	96,64	96,50	97,60	94,18	96,70	2
32768	96,0	97,44	96,78	97,80	96,36	98,20	1
65536	98,95	98,30	96,75	98,66	99,10	98,88	0,5

Fonte: Elaborado pelo autor

Tabela 7 – Resultados do experimento 2 - Teclado digital

Wavelet	F2 a B2	C3 a F#3	G3 a C#4	D4 a G#4	A4 a E5	F5 a C6
haar	42,03	39,25	37,80	37,12	35,0	34,90
sym8	76,82	75,40	72,30	72,0	69,25	70,0
sym16	77,10	79,20	75,22	79,68	72,25	74,90
coif3	43,05	39,0	44,22	38,95	36,38	36,0
coif5	46,88	46,50	48,55	44,90	40,55	40,10
beyl9	47,78	41,25	41,66	40,33	58,50	39,0
daub2	45,02	40,75	39,28	41,92	38,30	37,66
daub4	55,88	43,88	37,12	46,22	54,60	53,62
daub8	78,80	77,40	75,57	74,78	72,20	70,68
daub12	78,96	76,90	76,85	79,30	75,62	78,42
daub16	98,09	94,93	91,35	90,46	86,79	85,31
daub20	92,30	91,88	91,50	92,33	88,25	92,90

Fonte: Elaborado pelo autor

Cada evento musical foi executado e captado 250 vezes para o saxofone alto e 150 vezes para o teclado digital. O saxofone alto resultou no máximo de 237 padrões diferentes para um único evento, enquanto que o teclado digital resultou no máximo em 140 padrões, sendo este o motivo para a escolha do número total de janelas por evento para cada instrumento. O método foi capaz de analisar cinco janelas por segundo, permitindo classificar cinco eventos distintos por segundo.

As Tabelas 8 e 9 contêm um resumo das taxas de acertos utilizando a DB16, para intervalos de eventos musicais (coluna Evt) captados pelo saxofone alto e pelo teclado digital, respectivamente. A colunas Ac(%) e Err(%) representam a taxa de acertos e erros relativos à identificação da frequência de cada evento musical, enquanto que

Tabela 8 – Precisão de detecção de eventos usando a função DB16 - Saxofone alto

Evt	Ac(%)	Err(%)	Ac.8ª(%)	Err.8ª(%)	Evt	Ac(%)	Err(%)	Ac.8ª(%)	Err.8ª(%)
A#1	56,86	43,14	100,0	0,0	D3	100,0	0,0	96,07	3,93
B1	70,58	29,41	100,0	0,0	D#3	96,07	3,93	96,07	3,93
C2	96,07	3,93	66,66	33,34	E3	100,0	0,0	100,0	324
C#2	68,62	31,37	100,0	0,0	F3	70,58	29,42	100,0	0,0
D2	96,07	3,93	100,0	0,0	F#3	98,03	1,97	100,0	0,0
D#2	94,11	5,89	100,0	0,0	G3	96,07	3,93	88,23	11,77
E2	100,0	0,0	100,0	0,0	G#3	98,03	1,96	52,94	47,06
F2	78,43	21,57	100,0	0,0	A3	98,03	1,97	100,0	0,0
F#2	100,0	0,0	100,0	0,0	A#3	94,12	5,88	100,0	0,0
G2	50,0	50,0	66,66	33,34	B3	70,58	29,41	100,0	0,0
G#2	60,78	39,22	100,0	0,0	C4	94,11	5,89	100,0	0,0
A2	94,12	5,88	100,0	0,0	C#4	100,0	0,0	100,0	0,0
A#2	98,03	1,97	88,23	11,77	D4	92,15	7,85	100,0	0,0
B2	72,54	27,46	100,0	0,0	D#4	50,0	50,0	100,0	0,0
C3	90,19	9,81	82,35	17,65	E4	98,03	1,97	100,0	0,0
C#3	92,15	7,85	100,0	0,0	F4	96,07	3,93	96,07	3,93

Fonte: Elaborado pelo autor

as colunas Ac.8ª(%) e Err.8ª(%) representam a taxa de acertos e erros relativos as oitavas de cada evento. A Tabela 10 contém um resumo geral da acurácia na detecção de eventos para o saxofone alto e o teclado digital.

Em geral, o método proposto apresentou uma boa taxa de acertos, tanto nas baixas como nas altas frequências, mantendo uma média geral acima de 86% para taxa de acertos de eventos relativos à frequência fundamental e superior a 94% para taxa de acertos relativos a oitavas, como pode ser observado na Tabela 10.

Dos 76 eventos analisados, observando-se as taxas de acertos relativas à frequência fundamental, os eventos D#4 e G2 do saxofone alto resultaram na menor taxa (50%), enquanto que 52 eventos tiveram uma taxa maior que 80%. Para as taxas de acertos relativas às oitavas, o evento G#3 do saxofone alto proporcionou a menor taxa (52,64%), enquanto que 65 eventos tiveram uma taxa superior a 88%.

Na Figura 50, têm-se duas matrizes de confusão representando respectivamente a taxa de acertos por frequência fundamental e a taxa de acertos por oitavas de cada evento musical. As linhas na matriz representam os eventos previstos para o modelo, enquanto que as colunas representam os eventos identificados. Além disso, a diagonal principal representa os acertos da classificação. Cada matriz contém todos os eventos e_n tanto do saxofone alto quanto do teclado digital ($n = 76$), sendo a ultima linha/coluna

Tabela 9 – Precisão de detecção de eventos usando a função DB16 - Teclado digital

Evt	Ac(%)	Err(%)	Ac.8ª(%)	Err.8ª(%)	Evt	Ac(%)	Err(%)	Ac.8ª(%)	Err.8ª(%)
F2	97,43	2,57	100,0	0,0	D#4	74,07	25,93	100,0	0,0
F#2	100,0	0,0	66,66	33,34	E4	96,22	3,78	94,11	5,89
G2	91,78	8,22	70,58	29,42	F4	98,05	1,95	100,0	0,0
G#2	100,0	0,0	94,11	5,89	F#4	74,07	25,93	100,0	0,0
A2	98,44	1,56	100,0	0,0	G4	99,20	0,80	66,66	33,34
A#2	100,0	0,0	100,0	0,0	G#4	94,17	5,83	74,50	25,50
B2	99,03	0,97	100,0	0,0	A4	76,42	23,58	50,98	49,02
C3	96,05	3,95	100,0	0,0	A#4	74,07	25,93	100,0	0,0
C#3	100,0	0,0	100,0	0,0	B4	94,17	5,83	100,0	0,0
D3	100,0	0,0	100,0	0,0	C5	79,51	20,49	100,0	0,0
D#3	100,0	0,0	100,0	0,0	C#5	74,07	25,93	100,0	0,0
E3	98,80	1,20	100,0	0,0	D5	97,64	2,36	100,0	0,0
F3	95,65	4,35	88,23	11,77	D#5	100,0	0,0	100,0	0,0
F#3	74,07	25,93	100,0	0,0	E5	97,91	2,09	74,50	25,50
G3	96,59	3,41	66,66	33,34	F5	100	0,0	100,0	0,0
G#3	74,07	25,93	100,0	0,0	F#5	74,07	25,93	100,0	0,0
A3	98,85	1,15	100,0	0,0	G5	88,88	11,12	98,03	1,97
A#3	74,07	25,93	100,0	0,0	G#5	74,07	25,93	100,0	0,0
B3	97,14	2,86	100,0	0,0	A5	100	0,0	100,0	0,0
C4	98,73	1,27	100,0	0,0	A#5	74,07	25,93	100,0	0,0
C#4	100,0	0,0	100,0	0,0	B5	74,07	25,93	100,0	0,0
D4	97,46	2,54	100,0	0,0	C6	97,34	2,66	100,0	0,0

Fonte: Elaborado pelo autor

Tabela 10 – Média de resultados

Instrumento	Qtde eventos	J\evento	Acertos(%)	Erros(%)	Ac.8ª(%)	Err.8ª(%)
Saxofone alto	32	250	86,57	13,43	94,79	5,21
Teclado digital	44	150	87,54	12,46	94,20	5,80

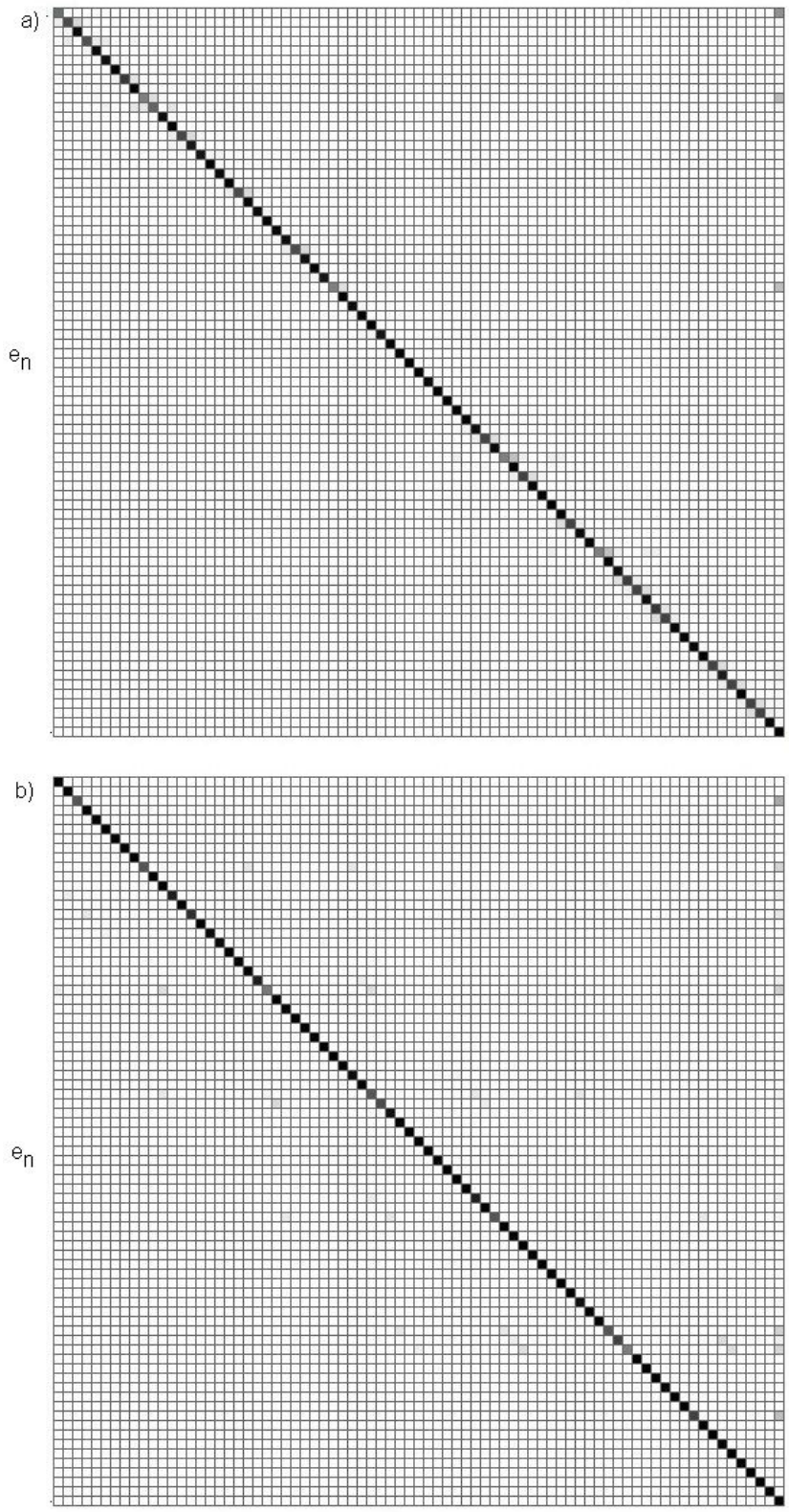
Fonte: Elaborado pelo autor

reservada para eventos não identificados. Observando a diagonal principal de cada matriz, pode-se notar que o sistema proporcionou uma boa taxa de acertos para cada evento musical, sendo que a taxa de erros, em sua maioria, situou-se próxima da diagonal principal e, em minoria, o restante foi classificado como “não identificado”.

5.1 Discussões

Neste trabalho foi proposto um método para classificação em tempo real de sinais musicais utilizando a Transformada *Wavelet-Packet*. O modelo foi desenvolvido com o objetivo de manter um compromisso entre eficiência computacional e acurácia, para eventos musicais originados de sons monofônicos captados por um teclado digital

Figura 50 – Matriz de confusão: (a) eventos e_n correspondentes a frequência das notas musicais e (b) eventos e_n correspondentes a oitavas de cada nota.



Fonte: Elaborado pelo autor

e um saxofone alto. Os eventos musicais identificados compreendem uma classe de sons elementares de uma partitura musical, condizentes ao aprendizado de músicos principiantes na prática e domínio básico do seu instrumento.

A análise *wavelet* mostrou-se útil na extração de características dos sinais musicais considerados, em razão da sua capacidade de analisar o sinal em diversos níveis de resolução. Ao comparar o desempenho de diferentes *wavelets*, nota-se que um esquema de decomposição utilizando a *wavelet* de Daubechies com suporte 16 cumpriu o requisito de manter uma boa localização frequencial, evitando redundâncias na representação de eventos musicais distintos. O procedimento envolvendo a aplicação da autocorrelação e de um limiar para seleção de picos de frequências sob o sinal transformado reduziu o número de harmônicos presentes e permitiu extrair diferentes padrões de conteúdos frequenciais.

Pode-se ainda constatar que a utilização do método proposto permitiu a classificação dos eventos com um baixo custo computacional, apresentando uma taxa de cinco eventos musicais identificados por segundo, a uma média geral acima de 86% para a taxa de acertos relativos à frequência fundamental, e 94% para as oitavas.

6 CONCLUSÃO

Neste trabalho, foi proposto um método capaz de reconhecer e classificar sinais musicais em tempo real, buscando manter um compromisso entre acurácia e baixa ordem de complexidade de tempo. A partir dos testes realizados, verificou-se que a análise *wavelet* apresentou resultados coerentes na etapa de extração de características dos sinais musicais, em função da sua habilidade para análise em níveis de resolução distintos. Particularmente, concluiu-se que um esquema de decomposição *wavelet-packet* com uma janela retangular de 8192 amostras utilizando os filtros digitais de Daubechies com suporte 16, eliminou redundâncias na representação tempo-frequência para eventos musicais distintos. O procedimento envolvendo a aplicação da autocorrelação e de um limiar para seleção de picos de frequências nos sinais transformados, reduziu o número de harmônicos presentes e permitiu extrair e classificar diferentes padrões com uma acurácia satisfatória para centenas de sinais, considerando-se uma rede neural do tipo RBF como a entidade inteligente.

Registre-se que a revisão bibliográfica realizada foi de considerável importância para a escolha do método apropriado de classificação dos eventos musicais. O desenvolvimento contou com dificuldades, dentre as quais a mais marcante foi a eliminação das redundâncias retromencionadas, oriundas das componentes harmônicas contidas em formas de onda diferentes daquelas puramente senoidais.

A etapa de testes, na qual foi concomitante com o desenvolvimento, cada uma das sub-partes do trabalho foi validada e aprimorada. Assim, é possível afirmar que o método proposto é promissor, tendo-se alcançado os objetivos. De uma forma geral, os resultados obtidos permitiram concluir que a metodologia proposta é viável, onde para os eventos musicais analisados (frequências das notas, eventos ligados e conjuntos formando fraseados), a família *wavelet* Daubechies apresentou melhores resultados em relação às demais wavelets analisadas. Destaque-se que o algoritmo proposto pode ser aplicado em projetos que envolvam o auxílio na aprendizagem prática instrumental para músicos iniciantes.

Como proposta para trabalhos futuros, buscar-se-á classificar outros eventos musicais, considerados particulares para determinadas categorias de instrumentos,

como por exemplo, um vibrato em um violino ou em uma flauta. Pretende-se, ainda, explorar questões relativas à recuperação de informações em sinais musicais.

REFERÊNCIAS

ADDISON, P.; WALKER, J.; GUIDO, R. Time-frequency analysis of biosignals. *IEEE Engineering in Medicine and Biology Magazine*, IEEE, v. 28, n. 5, p. 14–29, 2009.

ALVES, L. **Teoria Musical**. Vila Mariana: Irmãos Vitale, 2004.

AUGER, F. et al. Time-frequency toolbox. *CNRS France-Rice University*, Rice, p. 46, 1996.

BALDISSERA, F.; ORTH, A.; STEMMER, M. Análise da transformada de wavelet aplicada ao processamento frequencial de sinais. *SCPDI 2001 Simpósio Catarinense de Processamento Digital de Imagens*, Florianópolis, p. 13, 2001.

BARRETO, A. Perspectivas da ciência da informação. *Revista de Biblioteconomia de Brasília*, Brasília, v. 21, n. 2, p. 156–166, 1997.

BIANCHI, A. J. **Processamento de Áudio em Tempo Real em Dispositivos Computacionais de Alta Disponibilidade e Baixo Custo**. Tese (Doutorado) — Universidade de São Paulo, São Paulo, 2014.

BISHOP, C. **Neural Networks for Pattern Recognition**. Oxford: Oxford University Press, 1995.

BRAGA, A.; CARVALHO, A.; LUDERMIR, T. **Redes Neurais Artificiais:** *Teoria e Aplicações*. Rio de Janeiro: Livros Técnicos e Científicos, 2000.

BROWN, J.; PUCKETTE, M. An efficient algorithm for the calculation of a constant q transform. *The Journal of the Acoustical Society of America*, Acoustical Society of America, v. 92, n. 5, p. 2698–2701, 1992. Acesso em: 15 jan. 2017. Disponível em: <https://doi.org/10.1121/1.404385>.

CARDOSO, I. Ver uma melodia Sob o Prisma da Ciência. *ComCiência*, SciELO, p. 0 – 0, 00 2010. ISSN 1519-7654. Acesso em: 10 jan. 2017. Disponível em: <http://comciencia.scielo.br/scielo.php?script=sci_arttext&pid=S1519-76542010000200002&nrm=iso>.

CARVALHO, P. et al. Notas em matemática aplicada. *Sociedade Brasileira de Matemática Aplicada e Computacional*, São Carlos, v. 38, 2009.

CHEN, Y.; SHEN, H.; HSU, C. Fundamental frequency analysis on a harmonic power signal using fourier series and zero crossing algorithms. *Journal of Information Hidding and Multimedia Signal Processing*, v. 6, n. 5, 2015.

CRANITCH, M.; CYCHOWSKI, M.; FITZGERALD, D. Towards an inverse constant q transform. In: AUDIO ENGINEERING SOCIETY. *Audio Engineering Society Convention 120*. Paris, 2006.

CUNHA, G.; MARTINS, M. Tecnologia, produção & educação musical: Descompassos e desafinos. In: *Anais do Congresso RIBIE n. IV*. Brasília: [s.n.], 1998.

DAUBECHIES, I.; PAUL, T. Wavelets and applications. In: *8th International Congress of Mathematical Physics*. Marseille: [s.n.], 1987.

DINIZ, P.; SILVA, E.; NETTO, S. **Processamento Digital de Sinais: Projeto e Análise de Sistemas**. Porto Alegre: Bookman, 2014.

DONOHO, D.; JOHNSTONE, I. Ideal spatial adaptation by wavelet shrinkage. *Mathematical Reviews (MathSciNet): MR1311089 Zentralblatt MATH*, Biometrika, Oxford, v. 81, p. 425–455, 1994. Acesso em: 19 fev. 2017. Disponível em: <http://dx.doi.org/10.1093/biomet/81.3.425>.

ELFOULY, F. et al. Comparison between haar and daubechies wavelet transformations on {FPGA} technology. *International Journal of Computer, Information, and Systems Science, and Engineering*, Leicester, v. 2, n. 1, 2008.

ELOWSSON, A.; FRIBERG, A. Modelling perception of speed in music audio. *Forthcoming for Proc. of SMC*, Stockholm, 2013.

FARIA, R.; ZUFFO, J. Wavelets as a multiresolution analysis and synthesis technique for sound timbres edition. In: *Simpósio Brasileiro de Computação e Música, 2o/Congresso Da Sociedade Brasileira de Computação, 15o*. Canela: [s.n.], 1995. p. 198–204.

FARIAS, A. et al. Identificação de instrumentos musicais utilizando redes neurais artificiais. *Revista Liberato*, Novo Hamburgo, v. 10, n. 13, 2009.

FERNANDES, M.; NETO, A.; BEZERRA, J. Aplicação das redes RBF na deteção inteligente de sinais digitais. In: *IV Brazilian Conference on Neural Networks*. São José dos Campos: [s.n.], 1999. p. 226–230.

FERREIRA, S. **Sistema Especialista para Reconhecimento de Acordes Musicais em Tempo Real para Violão Elétrico Utilizando Técnicas de DSP**. Tese (Doutorado) — Universidade Federal da Bahia, Bahia, 2006.

FILLON, T.; PRADO, J. A flexible multi-resolution time-frequency analysis framework for audio signals. In: IEEE. *Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on*. [S.l.], 2012. p. 1124–1129.

GALVÃO, R. et al. Estudo comparativo sobre filtragem de sinais instrumentais usando transformadas de fourier e wavelet. *Química Nova*, SciELO Brasil, Recife, v. 24, n. 6, p. 874–884, 2001.

GANCHEV, T. et al. Wavelet basis selection for enhanced speech parametrization in speaker verification. *International Journal of Speech Technology*, Springer, v. 17, n. 1, p. 27–36, 2014.

GIANNAKOPOULOS, T.; PIKRAKIS, A. **Introduction to Audio Analysis:** *A Matlab Approach*. Cambridge: Academic Press, 2014.

GRIMALDI, M.; CUNNINGHAM, P.; KOKARAM, A. A wavelet packet representation of audio signals for music genre classification using different ensemble and feature selection techniques. In: ACM. *Proceedings of the 5th ACM SIGMM International Workshop on Multimedia Information Retrieval*. California, 2003. p. 102–108.

GUIDO, R. A note on a practical relationship between filter coefficients and scaling and wavelet functions of discrete wavelet transforms. *Applied Mathematics Letters*, Elsevier, v. 24, n. 7, p. 1257–1259, 2011.

GUIDO, R. Practical and useful tips on discrete wavelet transforms. *IEEE Signal Processing Magazine*, IEEE, v. 32, n. 3, p. 162–166, 2015.

GUIDO, R. A tutorial on signal energy and its applications. *Neurocomputing*, Elsevier, v. 179, p. 264–282, 2016.

GUIDO, R. Effectively interpreting discrete wavelet transformed signals [lecture notes]. *IEEE Signal Processing Magazine*, IEEE, v. 34, n. 3, p. 89–100, 2017.

HASSOUN, M. **Fundamentals of Artificial Neural Networks**. Cambridge: MIT Press, 1995.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. Porto Alegre: Bookman, 2001.

HAYKIN, S.; VEEN, B. **Sinais e Sistemas**. Porto Alegre: Bookman, 2001.

HU, J. Real-time perceptual tempo estimation for music signal based on envelope autocorrelation. In: IEEE. *Wireless Communications and Signal Processing (WCSP), 2010 International Conference on*. [S.l.], 2010. p. 1–4.

IAZZETTA, F. Interaçao, interfaces e instrumentos em música eletroacústica. In: PUC-RIO. *Proceedings of the II 'IHC-Interaç ao Humano-Computador Conference*. Campinas, 1998. p. 112–120.

INGALE, R. Harmonic analysis using FFT and STFT. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Sandy Bay, v. 7, n. 4, p. 345–362, 2014.

JENSEN, A.; COUR-HARBO, A. **Ripples in Mathematics:** *The Discrete Wavelet Transform*. Cambridge: Springer Science & Business Media, 2001.

JR, A. F.; DAMIANI, F. Extração automática de tempo musical utilizando transformada wavelet e o espectro rítmico. In: AES. *12th Congresso de Engenharia de Audio,18th Convenção Nacional da AES Brasil*. São Paulo, 2014. p. 9–16.

JUILLERAT, N.; ARISONA, S.; SCHUBIGER-BANZ, S. Enhancing the quality of audio transformations using the multi-scale short-time fourier transform. In: *Proceedings of the 10th IASTED International Conference*. Kailua-Kona: [s.n.], 2008. v. 623, p. 054.

JUNIOR, A. **Análise de Padrões Musicais Rítmicos e Melódicos Utilizando o Algoritmo de Predição por Correspondência Parcial**. Tese (Doutorado) — Universidade Federal da Paraíba, João Pessoa, 2011.

LATHI, B. **Sinais e Sistemas Lineares-2ed**. Porto Alegre: Bookman, 2007.

LEME, G.; BELLOCHIO, C. Professores de escolas de música: Um estudo sobre a utilização de tecnologias. *Revista da ABEM*, v. 15, n. 17, 2014. Acesso em: 30 jun. 2017. Disponível em: <http://www.abemeducacaomusical.com.br/revistas/revistaabem/index.php/revistaabem/article/view/284>.

LIMA, S. **Um Sistema para Transposição Automática de Seqüências MIDI Baseada em Alcance Vocal**. Tese (Doutorado) — Universidade Federal de Uberlândia, Uberlândia, 2006.

MALLAT, S. **A Wavelet Tour of Signal Processing**. Cambridge: Academic press, 1999.

MARCHI, E. et al. Audio onset detection: A wavelet packet based approach with recurrent neural networks. In: IEEE. *Neural Networks (IJCNN), 2014 International Joint Conference on*. [S.l.], 2014. p. 3585–3591.

MCLOUGHLIN, I. **Applied Speech and Audio Processing:** *with Matlab Examples*. Cambridge: Cambridge University Press, 2009.

MEDHI, B.; TALKUDHAR, P. Assamese vowel phoneme recognition using zero crossing rate and short-time energy. *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering*, Assam, v. 4, n. 4, 2014.

MENESES, E.; FORNARI, J. Educação musical através da improvisação livre com recursos computacionais: Contribuições e desafios. In: *Anais do XXV Congresso da Associação Nacional de Pesquisa e Pós-Graduação em Música - ANPPOM*. Vitória: [s.n.], 2015.

MILETTO, E. et al. Introdução à computação musical. In: *IV Congresso Brasileiro de Computação*. Porto Alegre: [s.n.], 2004.

MISITI, M. **Wavelet Toolbox:** *For Use with Matlab, User's Guide*. Massachusetts: MathWorks, 1997.

MOUTINHO, A.; NETO, L. Métodos de pré-processamento de sinais aplicados ao treinamento de redes neurais artificiais. In: *II Congresso Brasileiro de Computação, Universidade do Vale do Itajaí-Univale, Santa Catarina*. Vale do Itajaí: [s.n.], 2002.

MÜLLER, M. **Fundamentals of Music Processing:** *Audio, Analysis, Algorithms, Applications*. London: Springer, 2015.

NAGARAJ, K.; EVANS, B. Toward automatic transcription-pitch tracking in polyphonic environment. *Multidimensional Digital Signal Processing*, 2003. Acesso em: 30 jun. 2017. Disponível em: «http://www.ece.utexas.edu/~bevans/courses/ee381k/projects/spring03/nagaraj/LitSurveyReport.pdf>.

NIEVERGELT, Y. **Wavelets Made Easy**. Berlim: Springer, 1999. v. 174.

OLIVEIRA, H. Análise de fourier e wavelets: Sinais estacionários e não estacionários. *Recife: Ed. Universitária UFPE*, Recife, 2007.

OLIVEIRA, H. **Análise de Sinais para Engenheiros:** *Uma Abordagem via Wavelets*. Recife: Brasport, 2007.

OLIVEIRA, H.; FALK, T.; TÁVORA, R. Decomposição de wavelets sobre corpos finitos. *Journal of Communication and Information Systems*, v. 17, n. 1, p. 38–47, 2017.

POPESCU, A.; GAVAT, I.; DATCU, M. Wavelet analysis for audio signals with music classification applications. In: IEEE. *Speech Technology and Human-Computer Dialogue, 2009. SpeD'09. Proceedings of the 5-th Conference on*. [S.l.], 2009. p. 1–6.

POZZOLI, E. **Guia Teórico-Prático para o Ensino do Ditado Musical**. São Paulo: Ricordi, 1983.

PRIOLLI, M. **Teoria Musical:** *Princípios Básicos da Música para a Juventude*. Rio de Janeiro: Casa Oliveira de Música, 2015.

QUINTAIS, L. Cultura e cognição. *Coimbra: Biblioteca Mínima*, Coimbra, 2009.

ROQUE, T.; MENDES, R. Extração de descritores sonoros timbrísticos a partir da transformada wavelet packet. In: AES. *12th Congresso de Engenharia de Audio,18th Convençõ Nacional da AES Brasil*. São Paulo, 2014. p. 39–46.

ROZINAJ, G.; NAGY, M. An analysis/synthesis system of audio signal with utilization of an SN model. *Radioengineering*, Společnost Pro Radioelektronické Inženỳrství, Bratislava, 2004.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial:***Um Enfoque Moderno. 2. ed.* Berlim: Elsevier, 2004.

SHETE, D.; PATIL, S.; PATIL, S. Zero crossing rate and energy of the speech signal of devanagari script. *IOSR-JVSP*, v. 4, n. 1, p. 1–5, 2014.

SHIMAMURA, T.; KOBAYASHI, H. Weighted autocorrelation for pitch extraction of noisy speech. *IEEE Transactions on Speech and Audio Processing*, IEEE, v. 9, n. 7, p. 727–730, 2001.

SHIRADO, W. et al. Estudo comparativo entre algoritmos das transformadas discretas de fourier e wavelet. *Revista Brasileira de Computação Aplicada*, Passo Fundo, v. 7, n. 3, p. 97–107, 2015. [Online; Acesso em: 6 mar. 2017]. Disponível em: <http://www.seer.upf.br/index.php/rbca/article/view/4880/3505>.

SILVA, J.; CARVALHO, F.; MORET, M. Fourier e wavelets na transcrição musical do sinal de áudio. In: AES. *Anais do 4o Congresso Brasileiro de Engenharia de Audio da AES-Brasil*. São Paulo, 2006.

SOUZA, F. et al. Comparação das bases de wavelets ortonormais e biortogonais: Implementação, vantagens e desvantagens no posicionamento com gps. *Trends in Applied and Computational Mathematics*, Presidente Prudente, v. 8, n. 1, p. 149–158, 2007. Acesso em: 6 mar. 2017. Disponível em: <https://tema.sbmac.org.br/tema/article/view/242>.

SURAJ, A. et al. Discrete wavelet transform based image fusion and de-noising in {FPGA}. *Journal of Electrical Systems and Information Technology*, v. 1, n. 1, p. 72 – 81, 2014. ISSN 2314-7172. Disponível em: <http://www.sciencedirect.com/science/article/pii/S2314717214000075>.

SZCZUPAK, A.; BISCAINHO, L. Identificação de notas musicais em registros de violão solo. In: AES. *Anais do 7o Congresso Brasileiro de Engenharia de Audio da AES-Brasil*. São Paulo, 2009.

SZCZUPAK, A.; BISCAINHO, L.; CALÔBA, L. Identificação de notas musicais de violão utilizando redes neurais. In: AES. *Anais do 4o Congresso Brasileiro de Engenharia de Audio da AES-Brasil*. São Paulo, 2006.

TREVILLATO, N.; BARBEDO, J.; LOPES, A. Transcrição automática de sinais de áudio. In: *Anais do X Simpósio Brasileiro de Computação Musical*. Brasília: [s.n.], 2005. p. 291–294.